

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 12:15:08 ; Search time 17.4 seconds  
(without alignments)  
745.463 Million cell updates/sec

Title: US-09-810-506-2

Perfect score: 1816

Sequence: 1 MAPEINTKLTVPVHSATGGE.....FIEALSEAGALQYVKAPSAA 335

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	243	13.4	332	1	GLYG_MOUSE
2	241	13.3	349	1	GLYG_HUMAN
3	240	13.2	501	1	GYG2_HUMAN
4	235.5	13.0	332	1	GLYG_RABIT
5	231	12.7	332	1	GLYG_RAT
6	165	9.1	371	1	YAQC_SCHPO
7	129.5	7.1	380	1	GLG2_YEAST
8	122	6.7	618	1	GLG1_YEAST
9	110	6.1	286	1	GSPA_BACSU
10	105	5.8	1251	1	YMI1_YEAST
11	103.5	5.7	2157	1	POLG_HRV1B
12	96.5	5.3	338	1	RFAP_ECOLI
13	94.5	5.2	443	1	COBB_METJA
14	93.5	5.1	337	1	RFAP_SALTY
15	93.5	5.1	623	1	HCYE_EURCA
16	93	5.1	862	1	LOXA_PHAVU
17	92	5.1	630	1	TNPE_STAAU
18	92	5.1	2151	1	REPL_HANTV
19	91.5	5.0	336	1	RFAP_SALTY
20	91.5	5.0	342	1	SPDE_LYCES
21	91	5.0	315	1	SPD1_HYONI
22	90	5.0	372	1	GMD5_HUMAN
23	89.5	4.9	759	1	ARY2_CALVI
24	88	4.8	824	1	TGLK_RAT
25	87.5	4.8	324	1	CP21_HORVU
26	87	4.8	285	1	Y714_STRCO
27	87	4.8	766	1	DP44_HUMAN
28	86.5	4.8	570	1	NCAP_MOPEI
29	86	4.7	314	1	SPDE_NICSY
30	86	4.7	317	1	SPD2_DARTS
31	85.5	4.7	450	1	VD10_BPT5
32	85.5	4.7	630	1	TNPB_STAAU
33	85.5	4.7	697	1	YHFO_YEAST

#### ALIGNMENTS

RESULT 1  
GLYG\_MOUSE STANDARD; PRT: 332 AA.  
AC Q9R052;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Glycogenin-1 (EC 2.4.1.186).  
GN GYG OR GYG1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=20011289; PubMed=10542328;  
RA van Maanen M.H., Fournier P.A., Palmer T.N., Abraham L.J.;  
RT 'Characterization of mouse glycogenin-1 cDNA and promoter region.';  
RL Biochim. Biophys. Acta 1447:284-290(1999).  
CC -!- FUNCTION: SELF-GLYCOSYLATES, VIA AN INTER-SUBUNIT MECHANISM, TO  
CC FORM AN OLIGOSACCHARIDE PRIMER THAT SERVES AS SUBSTRATE FOR  
CC GLYCOGEN SYNTHASE.  
CC -!- CATALYTIC ACTIVITY: UDP-glucose + glycogenin = UDP +  
CC glucosylglycogenin.  
CC -!- COFACTOR: SELF-GLYCOSYLATION IS DEPENDENT ON THE PRESENCE OF  
CC DIVALENT METAL IONS OF WHICH MANGANESE ION IS THE MOST EFFECTIVE.  
CC -!- PATHWAY: GLYCOGEN BIOSYNTHESIS.  
CC -!- SUBUNIT: HOMODIMER, TIGHTLY COMPLEXED TO GLYCOGEN SYNTHASE.  
CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE, HEART, IN A LESSER EXTENT IN  
CC KIDNEY, LUNG AND BRAIN.  
CC -!- PTM: SELF-GLYCOSYLATED BY THE TRANSFER OF GLUCOSE RESIDUES FROM  
CC UDP-GLUCOSE TO ITSELF, FORMING AN ALPHA-1,4-GLYCAN OF AROUND 10  
CC RESIDUES ATTACHED TO TYR-194.  
CC -!- SIMILARITY: BELONGS TO THE GLYCOGENIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF114031; AAD48901.1; -  
CC MGD; MGI:1351614; Gyg1.  
CC InterPro: IPR002495; Glycosyl\_transf\_8.  
CC Pfam: PF01501; Glyco\_transf\_8; 1.  
CC Transferrase; Glycogen biosynthesis; Acetylation; Phosphorylation;  
CC Glycoprotein.  
CC INIT\_MET 0 0 BY SIMILARITY.  
CC ACT\_SITE 85 85 BY SIMILARITY.  
CC MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
CC MOD\_RES 43 43 PHOSPHORYLATION (BY CAPK) (BY  
CC SIMILARITY).  
CC CARBOHYD 194 194 O-LINKED (GLC . .) (BY SIMILARITY).









OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96069371; PubMed=8524228;  
 RA Cheng C., Mu J., Farkas I., Huang D., Goebel M.G., Roach P.J.;  
 RT "Requirement of the self-glucosylating initiator proteins Gligp and  
 RL Glig2p for glycogen accumulation in Saccharomyces cerevisiae.";  
 RL Mol. Cell. Biol. 15:6632-6640(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=S288C / FY1679;  
 RA Katsoulou C., Tzermia M., Tavernarakis N., Alexandraki D.;  
 RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast  
 RT chromosome X reveals 14 known genes and 13 new open reading frames  
 RT including homologues of genes clustered on the right arm of  
 RT chromosome XI.";  
 RL Yeast 12:787-797(1996).  
 CC -!- FUNCTION: SELF-GLUCOSYLATING INITIATOR OF GLYCOGEN SYNTHESIS. ACT  
 CC AS A PRIMER FOR THE ELONGATION REACTION CATALYZED BY GLYCOGEN  
 CC SYNTHASE.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOGENIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U25436; AAA91644.1; -;  
 DR EMBL; X87371; CAA60818.1; -;  
 DR EMBL; Z49412; CAA89432.1; -;  
 DR SGD; S0003673; GLG2.  
 KW Glycogen biosynthesis.  
 SQ SEQUENCE 380 AA; 44546 MW; 36BDF556DEF397C0 CRC64;

Query Match 7.1%; Score 129.5; DB 1; Length 380;  
 Best Local Similarity 22.0%; Pred. No. 0.00077;  
 Matches 66; Conservative 50; Mismatches 85; Indels 99; Gaps 17;

QY 21 KRAYVTFAGTGDYKGVGVGLAKGLRAKSKYPLVAVLPDV-----PEDHR 67  
 DB 3 KVAICTLLYSRDLPGALFLAYLOKL----LKHAVEDEITCLLIEKFLGDEFKP 57  
 QY 68 KQLVDQGVCKE---IEPVPPENQTEFAMAYVI-----NYSKLRIWEFVEYNKMI 116  
 DB 58 QEIALIRSLFKEITIIIEPLKQDKSEIKNKANLELLKRPESHTLLKARLWELVQFDQVL 117  
 QY 117 YLDGDIQVFDNIDHFLDLP-NGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTWPEAK 175  
 DB 118 FLDA-----TLPLNKEFEILRLYPEQT-----RFQIA---AVPD-IGWPD-- 155  
 QY 176 LGPKPPPLYFNAGMFVYEPNL-----STYHNLETVKIVPPTLFAEQDFLNNYF-----K 224  
 DB 156 -----MFTGVLLIPDLMATSLQDFLIKTVSIDG-----ADQIFNQFFNPICNYSK 204  
 QY 225 DIYKPIPPV-----YNLVL-----AMLWRHPENIELQVQVHYCAAGAKPW 266  
 DB 205 EVLHKVSPLMIEWIRLPETYNTVMPNYGYQSSPAM-----NFFQOHIRLTHFTGT-FKPW 257

RESULT 8  
 ID GLG1\_YEAST  
 AC P36143;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Glycogen synthesis initiator protein GLG1.  
 GN GLG1 OR YKR058W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96069371; PubMed=8524228;  
 RA Cheng C., Mu J., Farkas I., Huang D., Goebel M.G., Roach P.J.;  
 RT "Requirement of the self-glucosylating initiator proteins Gligp and  
 RL Glig2p for glycogen accumulation in Saccharomyces cerevisiae.";  
 RL Mol. Cell. Biol. 15:6632-6640(1995).  
 RN [2]  
 RP SEQUENCE OF 139-480 FROM N.A.  
 RX STRAIN=S288C;  
 RA van Vliet-Reedijk J.C., Planta R.J.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: SELF-GLUCOSYLATING INITIATOR OF GLYCOGEN SYNTHESIS. ACT  
 CC AS A PRIMER FOR THE ELONGATION REACTION CATALYZED BY GLYCOGEN  
 CC SYNTHASE.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOGENIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U25446; AAA91646.1; -;  
 DR EMBL; Z28283; CAA82136.1; -;  
 DR PIR; S38134; S38134.  
 DR SGD; S0001766; GLG1.  
 DR InterPro: IPR002495; Glycosyl\_transf\_8.  
 DR Pfam: PF01501; Glyco\_transf\_8; 1.  
 KW Glycogen biosynthesis.  
 SQ SEQUENCE 618 AA; 69912 MW; E5422F6ACEDD9CDA CRC64;

Query Match 6.7%; Score 122; DB 1; Length 618;  
 Best Local Similarity 18.5%; Pred. No. 0.0063;  
 Matches 71; Conservative 56; Mismatches 122; Indels 134; Gaps 18;

QY 18 GGEKRAYVTFAGTGDYKGVGVGLAKGLRK-----47  
 DB 2 GMYKKLAIATLLYSADYLPGVFALGHQVKNKLEAEAGKKGDIETCLIVTTSFLNGLTSELA 61  
 QY 48 ---AKSKYPLVAVLPDPVPEHRKQLVDQGVCKEIPVPEPNOTEFAMAYVINYNSKL 104  
 DB 62 KNILOSIYTKIVLVEPLNQCQESIQNSKNLALLE-----RPESFALI-----KA 107  
 QY 105 RIWEFVEYNKMIYLDGDIQVFDNIDHFLDLP-NGQFYAVMDCFCCKTWSHSPQYKIGYCCQ 163  
 DB 108 RLWELTQFQVLYLDS-----TLPLNKEFLKFLDIMSQTTS-----QVGAIA 151  
 QY 164 QCPDKVTWPEAKLGPKPPPLYFNAGMFVYEPNLSTYHNLETVKIVPPTLF-----AEQ 216  
 DB 152 D-----IGWPD-----MFTGVLLIPD-----ADTASVLYQNTFNTSTGSDQ 191  
 QY 217 DFLNNYF-----KDIYK-----PIPPVYNLVLMLWRHPE---NIELDQVQVHYC 259  
 DB 192 GILNQFFNOCCTDELVDKDSFREWQLSFTYNTVINLPNGYQSSPAMVFKPSIKLIHFI 251  
 QY 260 AAGAKPWFRTGEEENMDREDIKMLVKKWDIYND--ESLDYKNVY-----IGDSHK----- 308  
 DB 252 GK-HKPSLSWSQKNFIKNE---YHDQWNEVEEFKEEHQNLNNEVSKPKISDSKDTETPE 306  
 QY 309 -----KQOTLOOFTEALS 321  
 DB 307 TITPDAPPSPNEPTTNOEIDTIS 329







Tt		01-AUG-1992 (Rel. 23, Created)
Dt		01-OCT-1994 (Rel. 30, Last sequence update)
Dt		16-OCT-2001 (Rel. 40, Last annotation update)
Gn	Lipopolysaccharide I,2-glucosyltransferase [EC 2.4.1.58].	
Dc	RFAJ OR WAAJ OR B3626.	
De	Escherichia coli.	
Os	Bacteri; Proteobacteria; gamma subdivision; Enterobacteriaceae;	
Oc	Escherichia.	
Ox	NCBI_TaxId=562;	
Xl	[1]	
Xr	SEQUENCE FROM N.A.	
Rc	STRAIN=K12.	
Rx	MEDLINE=92325066; PubMed=1624461;	
Rt	Pradai E., Parker C.T., Schnaitman C.A.;	
Ra	"Structures of the rfaB, rfaI, rfaJ, and rfas genes of Escherichia	
Rg	coli K-12 and their roles in assembly of the lipopolysaccharide	
Rt	core.";	
Rt	J. Bacteriol. 174:4736-4745(1992).	
Rl	[2]	
Rn	SEQUENCE FROM N.A.	
Rp	STRAIN=K12 / MG1655;	
Rc	MEDLINE=94316500; PubMed=8041620;	
Rx	Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;	
Rt	"Analysis of the Escherichia coli genome. V. DNA sequence of the	
Ra	region from 76.0 to 81.5 minutes."	
Rg	Nucleic Acids Res. 22:2576-2586(1994).	
Rt	-I- CATALYTIC ACTIVITY: UDP-glucose + lipopolysaccharide = UDP + D-	
Cc	glucosyl-lipopolysaccharide.	
Cc	-I- PATHWAY: LIPOLYSACCHARIDE CORE BIOSYNTHESIS.	
Cc	-I- SIMILARITY: TO S.TYMPHURIUM RFAJ AND TO RFAI.	
Cc	-----	
Cc	This SWISS-PROT entry is copyright. It is produced through a collaboration	
Cc	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
Cc	the European Bioinformatics Institute. There are no restrictions on its	
Cc	use by non-profit institutions as long as its content is in no way	
Cc	modified and this statement is not removed. Usage by and for commercial	
Cc	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
Cc	or send an email to licensee@isb-sib.ch).	
Cc	-----	
Dr	EMBL; M80059; AAA24087.1; .	
Dr	EMBL; U00039; CAB34636.1; .	
Dr	EMBL; AE000440; AAC76650.1; .	
Dr	PfR; D42982; D42982.	
Dr	EcoGene; EG11353; rfaJ.	
Dr	InterPro; IPRO02495; Glycosyl_transf_8.	
Dr	Fram; PF01501; Glyco_transf_8; 1.	
Kw	Lipopolysaccharide biosynthesis; Glycosyltransferase; Transferase;	
Kw	Complete proteome.	
Ft	CONFLICT 310 338 RYKHLLVQHVIISGIITAGVCYLCKRYRK -> DINIF	
Ft	(IN REF. 1)	
Sq	SEQUENCE 338 AA; 39040 MW; 819428EA13F1959A CRC64;	
	Query Match 5.3%; Score 96.5; DB 1; Length 338;	
	Best Local Similarity 21.5%; Pred. No. 0.49;	
	Matches 54; Conservative 43; Mismatches 91; Indels 63; Gaps 13;	
Qy	89 QTEFAMAYVINYSKL-----RTWFEVEY-----NKMIYLDGDIQVFNDID 129   :                     :     :         :   :	
Db	81 QNLRLTYLRINTDKLOCLPCTQVWSRAMYFRFLFAQLGLTLDRLLYLDADVCKGDIS 140   :                     :     :         :   :	
Qy	130 HLFDLP-NGOFYAV--MDCEKCTWHSPOYKIGYCQCDDPKWTPEAKLGPKPPLYFN 185                                 :	
Db	141 QLLHLGLUNGAAVVVKDVPENMQEKAIVS-----RUSDPEL-LQG----YFN 180                                 :	
Qy	186 AGMEFYENPLSTSHNLTETKVIVPTTL-----FAEQDFLMNYFKDIYKPPIPVYNLV 238   :                     :     :         :   :	
Db	181 SGV-VY-LDLKKWADAKLTEKALSILMSKDKNVYKYPDOVMNVLLKGMTFLPREYNTIIY 238   :                     :     :         :   :	
Qy	239 AM-----LWRHPENIELDQVKVVHYCAAGAKPWRFTEEENNREDIKMLVKKKWDIY 291   :                     :     :         :   :	
Db	239 TIKSELKDOKTKYKKLTISTELIHVTGA-TKPHWKWAIYPSKYKYKIALSNSPWKDV-- 295   :                     :     :         :   :	

[illegible]

```

Db 289 ADI-----LYIGGYPELFKEEJSRNKEMIESKEFDGYIYGCGL-MYITKSDNVP 342
Qy 233 VYNI-VLAMLWRHPENIELDQVKVHVHYCAAGAKPWRFTGEE-----ENNDRDREIKMLVK 285
Db 343 VGLLNCNAVTKHVQGLSYKAEFELEDCLLGRKLPKGFHEPHYSKLVNKEERFAYKIE 402
Qy 286 KWDIYND-ESLOYKNVIGDSHKQOTLQOQTEALSEAG 324
Db 403 RGRGIINLDGIFNGKVLGALYLNHAVANPYFASSMVNFG 442

RESULT 14
RFAL_SALTY
ID RFAL_SALTY STANDARD; PRT; 337 AA.
AC P19816; O68267;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Lipopolysaccharide 1,3-galactosyltransferase (EC 2.4.1.44).
GN RFAL OR WAAI OR STM3718.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=98204873; PubMed=9535865;
RA Heinrichs D.E., Monteiro M.A., Perry M.B., Whitfield C.;
RT "The assembly system for the lipopolysaccharide R2 core-type of
RT Escherichia coli is a hybrid of those found in Escherichia coli K-12
RT and Salmonella enterica. Structure and function of the R2 Waak and
RT Waal homologs."
RL J. Biol. Chem. 273:8849-8859(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT Lt2."
RL Nature 413:852-856(2001).
RN [3]
RP PRELIMINARY SEQUENCE OF 1-197 FROM N.A.
RA Flock J.-I.;
RL Submitted (JUL-1990) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP PRELIMINARY SEQUENCE OF 198-337 FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=91045080; PubMed=2235496;
RA Carstentus P., Flock J.-I., Lindberg A.;
RT "Nucleotide sequence of rfai and rfaj genes encoding
RT lipopolysaccharide glycosyl transferases from Salmonella
RT typhimurium."
RL Nucleic Acids Res. 18:6128-6128(1990).
CC -1- FUNCTION: ADDS THE GALACTOSE(1) GROUP ON THE GLUCOSE(1) GROUP
CC OF LPS.
CC -1- CATALYTIC ACTIVITY: UDP-galactose + lipopolysaccharide = UDP + D-
CC galactosyl-lipopolysaccharide.
CC -1- PATHWAY: LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS.
CC -1- SIMILARITY: TO E.COLI RFAL AND TO RFAL.
CC -1- CAUTION: REF.3 AND REF.4 SEQUENCES DIFFER FROM THAT SHOWN DUE TO
CC FRAMESHIFTS AND OTHER SEQUENCING ERRORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

---

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF026386; AAC16412.1; -
DR EMBL; AE008873; AAL22577.1; -
DR EMBL; X53847; CAA37841.1; ALT_FRAME.
DR PIR; S12097; S12097.
DR StyGene; SG10337; rfai.
DR InterPro; IPR002495; Glycosyl_transf_8.
DR Pfam; PF01501; Glyco_transf_8; 1.
KW Lipopolysaccharide biosynthesis; Glycosyltransferase; Transferase;
FT Complete proteome.
FT CONFLICT 125 125 R -> RE (IN REF. 1).
FT CONFLICT 210 210 E -> Q (IN REF. 1).
SQ SEQUENCE 337 AA; 38905 MW; 6DB1FA1531AE25F7 CRC64;

Query Match 5.1%; Score 93.5; DB 1; Length 337;
Best Local Similarity 20.3%; Pred. No. 0.88;
Matches 47; Conservative 36; Mismatches 83; Indels 65; Gaps 11;

Qy 84 YPENQTFE-----AMAYVYVINYSKLRI-----WEFVEY-----NKM 115
Db 67 FDSQQRFEALAKQYATQIVYVILIDCERLKLSPSTKNWTATYFRFTIADYFSDKTRDV 126
Qy 116 IYLDGDIQVFNIDHLDL--PNGQFVAVMDFCEKWTWSHPQYKIGYCCQCPDKVTWPE 173
Db 127 LYLDADIACKGSIQELIDLNAENEIAAVV-AGELEW-----WTK 166
Qy 174 AKLGPKPP-----LYPNAGMFVYENLSTYHN-----LLETVKIVPPTLFAEQDFLNMY 222
Db 167 RSVSLATPCLVSGVFNAGFILINPLTAENISKKAEMLKDPVGVORITHLDQDVLNIF 226
Qy 223 PKDIYKPPVYVNLVLAHLWRHPENI--ELDQVKV-VHYCAAGAKPWRFTG 270
Db 227 LVNKARFVDDKKFTQFSLNVELKDSVINPVDATFVHYIGP-TKPHWSWG 276

RESULT 15
HCYE_EURCA
ID HCYE_EURCA STANDARD; PRT; 623 AA.
AC P02242;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hemocyanin E chain (HCE).
GN HCE.
OS Euryvelma californica (American tarantula).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Mygalomorphae; Theraphosidae; Aphonopelma.
OX NCBI_TaxID=29932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90319102; PubMed=2371273;
RA Voll W., Voit R.;
RT "Characterization of the gene encoding the hemocyanin subunit e from
RT the tarantula Euryvelma californicum."
RL Proc. Natl. Acad. Sci. U.S.A. 87:5312-5316(1990).
RN [2]
RP SEQUENCE.
RX MEDLINE=84059635; PubMed=6357986;
RA Schneider H.-J., Drexel R., Feldmaier G., Linzen B., Lottspeich F.,
RA Henschen A.;
RT "Hemocyanins in Spiders, XVIII. Complete amino-acid sequence of
RT subunit e from Euryvelma californicum hemocyanin."
RL Hoppe-Seyler's Z. Physiol. Chem. 364:1357-1381(1983).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91060544; PubMed=2246235;
RA Voit R., Feldmaier-Fuchs G.;
RT "Arthropod hemocyanins. Molecular cloning and sequencing of cDNAs
RT encoding the tarantula hemocyanin subunits a and e."

```

Search completed: July 2, 2002, 12:15:10  
Job time: 372 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 12:14:44 ; Search time 48.67 Seconds  
(without alignments)  
1190.740 Million cell updates/sec

Title: US-09-810-506-2  
Perfect score: 1816  
Sequence: 1 MAPEINTKLTVPVHSATGCE.....FTEALSEAGALQYKVPASAA 335

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1816	100.0	335	10	Q9FBX2 arabidopsis
2	1516.5	83.5	334	10	O80518 arabidopsis
3	1390.5	76.6	344	10	O22893 arabidopsis
4	1388	76.4	341	10	Q9XBJ7 brassica na
5	1345.5	74.1	334	10	Q9XGQ4 pisum sativ
6	1319	72.6	328	10	Q40710 oryza sativ
7	1298.5	71.5	334	10	O22693 arabidopsis
8	1284.5	70.7	333	10	Q9XGN4 ajuga rept
9	1219	67.1	221	10	Q945L1 arabidopsis
10	1216.5	67.0	292	10	Q9XGN3 ajuga rept
11	1198.5	66.0	361	10	Q9STQ9 arabidopsis
12	1198	66.0	333	10	Q9FFA1 arabidopsis
13	1110	61.1	345	10	O80766 arabidopsis
14	1053.5	58.0	318	10	Q947G8 lycopersico
15	763.5	42.0	213	10	Q9SPE1 vitis ripar
16	258.5	14.2	1201	10	O80649 arabidopsis

17	248	13.7	307	5	Q9W2J6
18	248	13.7	655	10	Q940B5 arabidopsis
19	248	13.7	659	10	Q9LSB1 arabidopsis
20	232	12.8	557	10	Q9FZ37 arabidopsis
21	225.5	12.4	372	3	O43061
22	216	11.9	614	10	Q94HG3
23	205.5	11.3	260	2	O53234
24	205	11.3	333	10	Q9S2B0
25	197	10.8	429	5	Q22997
26	194	10.7	284	5	Q9S050
27	194	10.7	300	5	O45819
28	189	10.4	291	16	Q9L856
29	186.5	10.3	283	12	Q9J879
30	181	10.0	276	12	Q99GV2
31	178	9.8	276	12	Q9E236
32	166	9.1	269	12	Q91F08
33	158.5	8.7	277	12	Q9P200
34	157.5	8.7	404	16	Q97P73
35	157.5	8.7	546	10	O04031
36	147.5	8.1	442	10	O23503
37	140	7.7	460	3	Q9Y761
38	140	7.7	470	10	Q9ZQPA
39	135.5	7.5	289	12	Q9E7F3
40	134	7.4	306	2	Q50948
41	134	7.4	307	2	Q50948
42	134	7.4	376	3	O43062
43	132	7.3	398	16	Q97P77
44	131	7.2	406	16	Q97P76
45	129.5	7.1	311	2	P96945

ALIGNMENTS

RESULT 1

ID	Q9FBX2	PRELIMINARY;	PRT;	335 AA.
AC	Q9FBX2;			
DT	01-MAR-2001 (TrEMBLrel. 16, Created)			
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	POTATIVE GALACTINOL SYNTHASE (AT1G56600/F25P12_16).			
GN	F25P12.95.			
OS	Arabidopsis thaliana (Mouse-ear cross).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,			
RA	Altati H., Nguyen M., Lam B., Southwick A., Bei, Buehler E., Chin C.,			
RA	Chiou J., Choi E., Dunn P., Gonzalez A., Howng B., Kim C., Koo T.,			
RA	Lee J.M., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharly N.,			
RA	Pham P., Sakano H., Schwartz J., Shinn P., Thaveri A., Toriumi M.,			
RA	Vaysberg M., Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;			
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,			
RA	Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,			
RA	Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,			
RA	Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,			
RA	Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,			
RA	Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,			
RA	Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,			
RA	Ecker J.R.;			
RT	"Arabidopsis ORF clones.";			
RN	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,			
RA	Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,			



RP	SEQUENCE FROM N.A.					
RA	Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,					
RA	Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,					
RA	Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,					
RA	Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,					
RA	Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,					
RA	Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,					
RA	Southwick A., Tracy S.E., Shinzato K., Davis R.W., Ecker J.R.,					
RA	Theologis A.;					
RT	"Full Length cDNA of gene t8I13.2/At2g47180 (GI:2275196).";					
RL	Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.					
DR	EMBL; AC003337; AAB63818.1; -;					
DR	EMBL; AY056139; AAL07218.1; -;					
DR	InterPro; IPR002495; Glycosyl_transf_8.					
DR	Pfam; PF01501; Glyco-transf_8; 1.					
SQ	SEQUENCE 344 AA; 39596 MW; 46784EA16BDD3A46 CRC64;					
	Query Match	76.6%;	Score 1390.5;	DB 10;	Length 344;	
	Best Local Similarity	74.7%;	Pred. No. 1.1e-121;			
	Matches 257;	Conservative 31;	Mismatches 47;	Indels 9;	Gaps	
QY	- 1 MAPEI-----NTKLTVPVHSATGGEKRAYVTFLAGTGDYVKGVGLAKLRKAISKY 52	:	:	:	:	:
Dd	1 MAPGLQTADAMSTVITIKPSLPVSQDSRAYVTFLAGNGDYVKGVGLAKLRKKVKSAY 60	:	:	:	:	:
QY	53 PLVVAVILPDPEDHRKQLVDGGCVKEIEFPVYPENOTEFMAYYVINYSKLRIWEFVEY 112	:	:	:	:	:
Dd	61 PLVVAVILPDPPEHREIRLLVDGGCIVREIEFPVYPENQTFAMAYYVINYSKLRIWKAFYE 120	:	:	:	:	:
QY	113 NKMIYLDGDTQVFNDIHDLFDLPNGOFYAVMDCFCSEKTWSHPQYKIGYCQCQCDPKVTP 172	:	:	:	:	:
Dd	121 SKMIYLDGDIQVYNIDHLFDLPDGYLYAVMDCFCSEKTWSHTPOYKIRYCQCQCDPKVQP 180	:	:	:	:	:
QY	173 EAKLGKPPPIYFNAGMFVTEPNLSYHNLETVKIVVPTPLFAQDFLNMYFKDIYKPDP 232	:	:	:	:	:
Dd	181 KABELGEPALLYFNAGMFLYEPNLETYEDLLRTLKITPTPFARQDFLNMYFKKIYKPIEL 240	:	:	:	:	:
QY	233 VYNVLVLAMLWRHPENTELDQKVYVHYCAAGAKPWRTGTEENNMDREDIKMLYKKWWDIYN 292	:	:	:	:	:
Dd	241 VYNVLVLAMLWRHPENVDELKVKVYVHYCAAGSKFWRTGTEANNEREDIKMLYKKWWDIYD 300	:	:	:	:	:
QY	293 DESLDYKN-VVIGDSHKQOOLQOFIEALSEAGALQYVRAPSAA 335	:	:	:	:	:
Dd	301 DESLDYKKPVTVVDVTLNKLKPFITALTEAGRLNYVTAPSA 344	:	:	:	:	:
RESULT	4					
ID	Q9XEJ7 PRELIMINARY;	PRT;	341 AA.			
AC	Q9XEJ7;					
DT	01-NOV-1999 (TrEMBLrel. 12, Created)					
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)					
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)					
DE	GALACTINOL SYNTHASE (FRAGMENT).					
GS	GS.					
OS	Brassica napus (Rape).					
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;					
OC	eurosid II; Brassicales; Brassicaceae; Brassica.					
ON	NCBI_TaxID=3708;					
RX	[1]					
SEQUENCE FROM N.A.						
RP	Georges F., Bock C.A., Hussain A.K., Yan W., Keller W.A.;					
RA	"Cloning and Characterization of a Full Length cDNA for Galactinol					
RT	Synthase from Brassica napus.";					
RL	Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.					
DR	EMBL; AF106954; AAD26116.1; -;					
DR	InterPro; IPR002495; Glycosyl_transf_8.					
DR	Pfam; PF01501; Glyco-transf_8; 1.					
FT	NON_TER 1					
SQ	SEQUENCE 341 AA; 39059 MW; 12ED74936583D157 CRC64;					

```
QY 121 DIQVFNIDHLDLPNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTPWPEAKLGP 180
Db 120 DIQYENIDHLDLPDGYFYAVMDCFCCKTWSHTPOYKIGYCCQCPKQWPK-EMGEPP 178
QY 181 PLYENAGMFVYEPNLSTYHNLETKIVPTLFAEQDFLNMFKDIYKPIPPVYNLVLAM 240
Db 179 SLFYNAGMFVEFSVEYIDLLTKCTAPTAPFADQDFLNMFKDIYRPIPLVYNLVLAM 238
QY 241 LWRHPENIELDQVVKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWMDIYNDESLDYKN 300
Db 239 LWRHPENVELRKVVVHYCAAGSKPWRTGKEENMQREDIKMLVQKWLDIYSSSLDYKK 298
QY 301 VVIGDSHKKQOTLQO-FIEALSAGALQYVKAPSA 335
Db 299 NLSNCETORNDVEEPVQALSEGVRYVYTPASAA 334

RESULT 6
Q40710 PRELIMINARY; PRT; 328 AA.
AC Q40710;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE WSI76 PROTEIN INDUCED BY WATER STRESS.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=SOMEWAKE;
RX MEDLINE=95036005; PubMed=7948880;
RA Takahashi R., Joshee N., Kitagawa Y.;
RT "Induction of chilling resistance by water stress, and cDNA sequence
RT analysis and expression of water stress-regulated genes in rice.";
RL Plant Mol. Biol. 26:339-352(1994).
DR EMBL; D26537; BAA05538.1; -
DR InterPro; IPR002495; Glycosyl_transf_8.
DR Pfam; PF01501; Glyco_transf_8; 1.
SQ SEQUENCE 328 AA; 37698 MW; D87DECDA7AF98B7 CRC64;

Query Match 72.6%; Score 1319; DB 10; Length 328;
Best Local Similarity 71.0%; Pred. No. 5.2e-115;
Matches 238; Conservative 41; Mismatches 48; Indels 8; Gaps 2;

QY 1 MAPEINTKLTVPVHSATGGEKRAYVTFLAGTDYKGVVGLAKGRKAKSKYPLVAVLP 60
Db 2 MGNVSSSE----KKALAAAKRRAYVTFLAGDGYKGVVGLAKGRVRSAYPLVAVLP 57
QY 61 DVPEDHRKQLVDGCVVKEIEPYPENOTEFAMAYVYINYSKLRIWFEVYENKMIYLDG 120
Db 58 DVPGEHRKLVEGCVVRIOYPIPPESOTQFAMAYVYINYSKLRIWFEVYENKMIYLDG 117
QY 121 DIQVFNIDHLDLPNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTPWPEAKLGP 180
Db 118 DIQVFNIDHLDLKGAFYAVKDCFCCKTWSHTPOYDYGCCQCPDEAVWPPERELGP 177
QY 181 PLYENAGMFVYEPNLSTYHNLETKIVPTLFAEQDFLNMFKDIYKPIPPVYNLVLAM 240
Db 178 PLYENAGMFVEHPGLGTAKDLDLALVPTPTFAEQDFLNMFFRQYKPIPPVYNLVLAM 237
QY 241 LWRHPENIELDQVVKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWMDIYNDESLDYKN 300
Db 238 LWRHPENVDLQVVKVHYCAAGSKPWRTGKEENMQREDIKMLVQKWLDIYSSSLDYKK 297
QY 301 VVIGDSHKKQOTLQO-FIEALSAGALQYVKAPSA 335
Db 298 ----EEDNADEASQPMRTALAEAGAVKYVFPASAA 328
```

```
RESULT 7
O22693 PRELIMINARY; PRT; 334 AA.
AC O22693;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F8A5.2 PROTEIN.
GN F8A5.2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,
RA Au M., Araujo R., Brendel V., Buehler E., Dewar K., Feng J., Kim C.,
RA Li Y., Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M.,
RA Vysotskaia V.S., Yu G., Ecker J., Theologis A., Davis R.W.;
RT "Genomic sequence of Arabidopsis.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC002292; AAB71970.1; -
DR InterPro; IPR002495; Glycosyl_transf_8.
DR Pfam; PF01501; Glyco_transf_8; 1.
SQ SEQUENCE 334 AA; 38511 MW; 591D3D441F01FA75 CRC64;

Query Match 71.5%; Score 1298.5; DB 10; Length 334;
Best Local Similarity 70.7%; Pred. No. 4.4e-113;
Matches 241; Conservative 35; Mismatches 52; Indels 13; Gaps 4;

QY 1 MAPEINTK---LTPVHSATGGEKRAYVTFLAGTDYKGVVGLAKGRKAKSKYPLVVA 57
Db 1 MAPEISVNPWVLSKHAQAP--PRAYVTFLAGDGYKGVVGLAKGRKAKSKYPLVVA 58
QY 58 VLPDVPEDHRKQLVDGCVVKEIEPYPENOTEFAMAYVYINYSKLRIWFEVYENKMIY 117
Db 59 MLPDPVEHREILRSQCVVREIEPVPDQVFEFAMAYVYINYSKLRIWFEVYENKMIY 118
QY 118 LDGDIQVFNIDHLDLPNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTPWPEAKLG 177
Db 119 LDADIQVFNIDHLDLPNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTPWPEAKLG 178
QY 178 KPPLYFNAGMFVYEPNLSTYHNLETKIVPTLFAEQDFLNMFKDIYKPIPPVYNLV 237
Db 179 PPPLYFNAGMFVEPSPLYESLQTLTPPSPPFAEQDFLNMFFKVKYKPIPLVYNLV 238
QY 238 LAMLRHPENIELDQVVKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWMDIYNDES 297
Db 239 LAMLRHPENVELEKVKVHYCAAGSKPWRTGEEANMDREDIKMLVKKWMDIYNDES 298
QY 298 YKNVVGDSHK---KQOTLQOFTIEALSAGALQYVKAPSA 335
Db 299 FSKIPADAETVTKSSILASVLEP-----EMTYFPAPSA 334

RESULT 8
Q9XGN4 PRELIMINARY; PRT; 333 AA.
AC Q9XGN4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GALACTINOL SYNTHASE, ISOFORM GOLS-1 (EC 2.4.1.123).
GN GOLS.
OS Ajuga reptans (bugle).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Lamiaceae; Ajuga.
OX NCBI_TaxID=38596;
RN [1]
```



```

RP SEQUENCE FROM N.A.
RX MEDLINE=20224020; PubMed=10758476;
RA Sprenger N., Keller F.;
RT "Allocation of raffinose family oligosaccharides to transport and
RT storage pools in Ajuga reptans: the roles of two distinct galactinol
RT synthases.";
RL Plant J. 21:249-258(2000).
DR EMBL: AJ237693; CAB51533.1; -.
DR InterPro: IPR002495; Glycosyl_transf_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 333 AA; 37881 MW; B3D575F9084DEFF6B CRC64;

Query Match 70.7%; Score 1284.5; DB 10; Length 333;
Best Local Similarity 73.0%; Pred. No. 8.8e-112;
Matches 233; Conservative 38; Mismatches 41; Indels 7; Gaps 4;

QY 19 GKKRAYVTFIAGTGVYGVVGLAKGLRKAQSKYPLVAVLPDVPEDHRKQLVDGCVVK 78
DB 20 GAKGYVTFIAGTGVYGVVGLAKGLRKAQSKYPLVAVLPDVPEDHRKQLVDGCVVK 79
QY 79 EIEPVYPENQTEFAMAYVINYSKRIWFEVYKMYLDGDIQVFDNDHLDLPNGO 138
DB 80 EIEPIYPPANQIFAMAYVINYSKRIWFEVYKMYLDADIQVFNIDHLDLPNGY 139
QY 139 FYAVMDFCEKTSWSPQYKIGYVQCCQCPDKVTPPEAKLGPPLYPNAGMFVYEPNLSY 198
DB 140 FYAVMDFCEKTSWSPQYKIGYVQCCQCPDKVTPP-AQMGSPPPLYPNAGMFVPEPSKTY 198
QY 199 HNLETKVIVPPTLFAEQDFLNMYFKDIKPIPPVNLVLAHLWRHPENIELDQVKKVHY 258
DB 199 QTLHLTRITPPPPFAEQDFLNMYFKDIKPIPPVNLVLAHLWRHPENIELDQVKKVHY 258
QY 259 CAAGAKPWRFTGEENMDREDIKMLYKKWVDYNDKYNVIGDGHKQQT--LQOF 316
DB 259 CAAGSKPWRFTGEENMDREDIKMLYKKWVDYNDKYNVIGDGHKQQT--LQOF 315
QY 317 IEALSEAGALQYVKAPSA 335
DB 316 IASLPEP-AVSYIPAPSA 333

RESULT 9
Q945L1 PRELIMINARY; PRT; 221 AA.
AC Q945L1;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE ATIG56600/475P12.16.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cdna clones.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF412094; AAL06547.1; -.
SQ SEQUENCE 221 AA; 25759 MW; 30D65575CB2D706C CRC64;

Query Match 67.1%; Score 1219; DB 10; Length 221;

```

```

Best Local Similarity 100.0%; Pred. No. 6.7e-106;
Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 MIYLDGDIQVFDNDHLDLPNGQFYAVMDFCEKTSWSPQYKIGYVQCCQCPDKVTPPEA 174
DB 1 MIYLDGDIQVFDNDHLDLPNGQFYAVMDFCEKTSWSPQYKIGYVQCCQCPDKVTPPEA 60
QY 175 KLGPKPPLYFNAGMFVYEPNLSYHNLETKVIVPPTLFAEQDFLNMYFKDIKPIPPVY 234
DB 61 KLGPKPPLYFNAGMFVYEPNLSYHNLETKVIVPPTLFAEQDFLNMYFKDIKPIPPVY 120
QY 235 NIVLAMLWRHPENIELDQVKKVHYCAAGAKPWRFTGEENMDREDIKMLYKKWVDYNDK 294
DB 121 NIVLAMLWRHPENIELDQVKKVHYCAAGAKPWRFTGEENMDREDIKMLYKKWVDYNDK 180
QY 295 SLDYKNVIGDGHKQQTLOQFIQFIALSEAGALQYVKAPSA 335
DB 181 SLDYKNVIGDGHKQQTLOQFIQFIALSEAGALQYVKAPSA 221

RESULT 10
Q9XGN3 PRELIMINARY; PRT; 292 AA.
AC Q9XGN3;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GALACTINOL SYNTHASE, ISOFORM GOLS-2 (EC 2.4.1.123) (FRAGMENT).
GN GOLS.
OS Ajuga reptans (bugle).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Lamiaceae; Ajuga.
OX NCBI_TaxID=38596;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20224020; PubMed=10758476;
RA Sprenger N., Keller F.;
RT "Allocation of raffinose family oligosaccharides to transport and
RT storage pools in Ajuga reptans: the roles of two distinct galactinol
RT synthases.";
RL Plant J. 21:249-258(2000).
DR EMBL: AJ237694; CAB51534.1; -.
DR InterPro: IPR002495; Glycosyl_transf_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
KW Transferase; Glycosyltransferase.
FT NON_TER 1
SQ SEQUENCE 292 AA; 33768 MW; 7297EBC07071C2B CRC64;

Query Match 67.0%; Score 1216.5; DB 10; Length 292;
Best Local Similarity 75.4%; Pred. No. 1.7e-105;
Matches 224; Conservative 29; Mismatches 39; Indels 5; Gaps 3;

QY 39 VGLAKGLRAKSKYPLVAVLPDVPEDHRKQLVDGCVVKEIEPYPPNQTEFAMAYV 98
DB 1 VGLAKGLRKVGTYIPLVAVLPDVPPEHRIILVEQGVVREIEPYPPNHTEFAMAYV 60
QY 99 INYSKLRIWFEVYKMYLDGDIQVFDNDHLDLPNGQFYAVMDFCEKTSWSPQYK 158
DB 61 INYSKLRIWFEVYKMYLDGDIQVFNIDHLDLPNGQFYAVMDFCEKTSWSPQYK 120
QY 159 IGYCQCCQCPDKVTPPEAKLGPPLYPNAGMFVYEPNLSYHNLETKVIVPPTLFAEQDF 218
DB 121 IGYCQCCQCPDKVTPPEAKLGPPLYPNAGMFVYEPNLSYHNLETKVIVPPTLFAEQDF 179
QY 219 LNMVFKDIKPIPPVNLVLAHLWRHPENIELDQVKKVHYCAAGAKPWRFTGEENMDRE 278
DB 180 LNMVFKDIKPIPPVNLVLAHLWRHPENIELDQVKKVHYCAAGAKPWRFTGEENMDRE 239
QY 279 DIKMLYKKWVDYNDKYNVIGDGHKQQTLOQFIQFIALSEAGALQYVKAPSA 335
DB 240 DIKMLYKKWVDYNDKYNVIGDGHKQQTLOQFIQFIALSEAGALQYVKAPSA 292

```



RA Liu S., Li J., Araujo R., Au M., Brendel V., Buehler E., Conway A.B.,  
RA Conway A.R., Dewar K., Feng J., Kim C., Kurtz D., Li Y., Palm C.J.,  
RA Shinn P., Sun H., Davis R.W., Ecker J.R., Federspiel N.A.,  
RA Theologis A.;  
RT "Arabidopsis thaliana chromosome 1 BAC T13D8 sequence";  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RC Theologis A.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC004473; AAC24075.1; -.  
DR InterPro: IPR002495; Glycosyl\_transf\_8.  
DR Pfam: PF01501; Glyco.transf\_8; 1.  
DR SOURCE 345 AA; 39372 MW; B4EDE0FDA075E69 CRC64;  
SD

Query Match	61.1%	Score 1110;	DB 10;	Length 345;
Best Local Similarity	59.9%	Pred. No. 1.9e-95;		
Matches 203: Conservative	46;	Mismatches 64;	Indels 26;	Gaps 3;

Qy	20	E	K	R	A	V	T	F	L	A	G	T	D	V	V	K	V	G	V	L	A	K	L	R	K	A	K	S	Y	P	L	V	A	V	L	P	D	V	P	E	D	H	R	Q	L	V	D	Q	C	V	V	K	E	79		
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
D6	10	K	R	A	V	T	F	L	A	G	N	D	V	V	K	V	G	V	L	A	K	L	R	K	V	K	S	A	P	L	V	A	M	L	P	D	V	P	E	E	H	R	E	L	L	R	S	O	C	I	V	R	E	69		

```

Qy 80 IEPVPPENQTEFAMAYVIVNSKLRIW-----EFVEYNKIYIL 118
    |||||::: : ||||| ||||| | |||||
Db 70 IEPVPPENQCARAAVIVNSKLTATNVSUVIVTPTHEVESIPISTLNFEFFYNKIYIL 129

```

[illegible]

QY 179 KPPLYFNAGMFVYENLSTYHNLLTETVKIVPTLFAEQDFLNMFKDIYKPIPPVYNLVL 238

```

00      FF0F1E980H V E N T L I C U S Q D Q Y I L A T T R M Q Z L A K H C O R T A N T P O R T I N E R E
239 AMUWRHPENTELDQVKKVVHYCAAGAKPWRFTGEENMDREDIKMLVKKWDIYNDESLDY 298
:::||||| :::: ||||| |:::||| ||||| :|||:|||||:|||||:

```

QY 299 KNVIGDSHKKQOTLQQ--FTIALSEAGALQYVKAPSA 335  
+ + . . . + . . . + . . . + . . .  
DD 250 SVLMKHPGRVDLESYKVWHICPPGSPFKWTIGEENFMDKEDVANMLIKAWDILNDESLDF

Db 310 KPKSPAD---LEATVLESTIITASVTEAPLSYSPAAPSAA 345

Q947G8  
ID Q947G8  
AC Q947G8; PRELIMINARY; PRT; 318 AA.

01-DEC-2001 (trEMBLrel. 19, Last sequence update)  
DT  
01-DEC-2001 (trEMBLrel. 19, Last sequence update)  
DT  
01-DEC-2001 (trEMBLrel. 19, Last annotation update)  
DT  
DE PUTATIVE GALACTINOL SYNTHASE 1 (EC 2.4.1.123).

GN GOLDS' I.  
OS *Lycopersicon esculentum* (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

QC Asteridae: euasterids I; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RD SEQUENCE FROM NCBI

RA Downie B., Dahal P., Nonogaki H., Gurusingham S., Yim K., Fukunaga K.,  
RA Alvarado V., Bradford K.J.;  
RT "Galactinol synthase gene expression in tomato seeds.";

DR EMBL; AF311943; AAL26804.1; -.  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 318 AA; 36402 MW; AF66D71F45097E52 CRC64;

Query Match 58.0%; Score 1053.5; DB 10; Length 318;  
Best Local Similarity 58.4%; Pred. No. 3.3e-90;

[illegible]

```

RESULT 15
Q9SPE1
ID Q9SPE1 PRELIMINARY; PRT; 213 AA.
AC Q9SPE1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE GALACTINOL SYNTHASE (FRAGMENT).
GN WSI76.
OS Vitis riparia (Frost grape) (Vitis vulpina).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
OC Vitis.
OX NCBI_TaxID=96939;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FLOWER BUDS;
RA Li X.-Z., McKersie B.D.;
RT "Freezing tolerance in grape flower buds.";
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF178569; AAD55726.1; -.
DR InterPro; IPR002495; Glycosyl_transf_8.
DR Pfam; PF01501; Glyco_transf_8; 1.
FT NON_TER 1
SO SEQUENCE 213 AA; 23590 MW; 6D008A67FF729EEA CRC64;

```

Query Match	42.0%;	Score 763.5;	DB 10;
Best Local Similarity	80.9%;	Pred. No. 2.4e-63;	
Matches 144:	Conservative	12: Mismatches	15: Indels
			7: Gaps

QY 15 SATGGEK-----RAYVTFAGTGDYVRGVVGLAKGLRKAQKAKSYPLVVAVLPDPVPEHRK 68

Db 5 SATCLAKAASSTSSPAYVTFAGNGDYGKGVVGLAKGLRKKVTAYPVVAVLPDPVPAEHR 64

QY	69	QLVDGCVVKEIEPVYPENOTEFAMAYVYN	SKLRIEWEFVEYNKMYLDGDIQVFDNI	128
SY	69	QLVDGCVVKEIEPVYPENOTEFAMAYVYN	SKLRIEWEFVEYNKMYLDGDIQVFDNI	124

[illegible]

1.25 DIME DEDDOLL INVINDICA CERAMONICA E ANTICO ORO E STAGNO. TERNI 1998-1999. 100

Search completed: July 2, 2002, 12:14:44

Tue Jul 2 14:24:07 2002

us-09-810-506-2.rspt

Page 8

Job time: 376 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 12:09:33 ; Search time 29.88 seconds  
(without alignments)  
1077.307 Million cell updates/sec

Title: US-09-810-506-2  
Perfect score: 1816  
Sequence: 1 MAPEINTKLTVPVHSATGGE.....FIEALSEAGALQYVKAPSAA 335

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1816	100.0	335	2 G96607	probable galactinol synthase
2	1516.5	83.5	334	2 F86226	hypothetical prote
3	1390.5	76.6	344	2 A84912	probable galactinol synthase
4	1319	72.6	328	2 T07610	WS176 protein - ri
5	1298.5	71.5	334	2 H96629	hypothetical prote
6	1198.5	66.0	361	2 T06009	hypothetical prote
7	1110	61.1	345	2 T02295	hypothetical prote
8	258.5	14.2	1201	2 T00444	hypothetical prote
9	241	13.3	333	1 JC4695	glycogenin glucosyl transferase
10	235.5	13.0	332	1 A45094	glycogenin glucosyl transferase
11	232	12.8	557	2 H96590	hypothetical prote
12	225.5	12.4	372	2 T40489	hypothetical prote
13	205.5	11.3	260	2 A45729	glycogenin homolog
14	205	11.3	333	2 T05984	hypothetical prote
15	197	10.8	429	2 T31001	hypothetical prote
16	194	10.7	300	2 T25275	hypothetical prote
17	165	9.1	371	2 S55490	hypothetical prote
18	157.5	8.7	404	2 G95206	glycosyl transferase
19	157.5	8.7	546	2 G86221	protein F7G19.14
20	147.5	8.1	442	2 A71433	hypothetical prote
21	143.5	7.9	321	2 T24773	hypothetical prote
22	140	7.7	470	2 A84772	probable glycogenin
23	134	7.4	376	2 T40488	hypothetical prote
24	132	7.3	398	2 G95205	glycosyl transferase
25	131	7.2	406	2 H95205	glycosyl transferase
26	129.5	7.1	380	2 S55176	GLG2 protein - yea
27	118	6.5	114	2 JC6077	PI3 protein - Leuc
28	110	6.1	286	2 S16423	general stress pro
29	110	6.1	491	2 S58330	probable membrane

## ALIGNMENTS

RESULT 1

G96607

probable galactinol synthase F25P12.95 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: G96607

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Lin, Y.; Liu, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719

A;Accession: G96607

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-335 <STO>

A;Cross-references: GB:AE005173; NID:g9954752; PIDN:AAG09103.1; GSPDB:GN00141

C;Genetics:

A;Gene: F25P12.95

A;Map position: 1

Query Match 100.0%; Score 1816; DB 2; Length 335;  
Best Local Similarity 100.0%; Pred. No. 4.2e-143;  
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPEINTKLTVPVHSATGGEKRAYVTFAGTGYVGVVGLAKGLKAKSKYPLVVAVLP 60

|||||  
Db 1 MAPEINTKLTVPVHSATGGEKRAYVTFAGTGYVGVVGLAKGLKAKSKYPLVVAVLP 60

|||||

QY 61 DVPEDEHRLQVDOGVVYKTEPEVYPPEQTAFAMAYVINYSLRIWFEVYKMYLDLG 120

|||||

Db 61 DVPEDEHRLQVDOGVVYKTEPEVYPPEQTAFAMAYVINYSLRIWFEVYKMYLDLG 120

|||||

QY 121 DIQVFNIDHFLDPLNGQFVAVMDCFEKTSWSPQYKIGYCCQCPDKVTPWEAKLGPKP 180

|||||

Db 121 DIQVFNIDHFLDPLNGQFVAVMDCFEKTSWSPQYKIGYCCQCPDKVTPWEAKLGPKP 180

|||||

QY 181 PLFVFNAGMFYFENLSTYHNLLTETVKIVPPTLFAEQDFLNMYPKDIYKPIPPVYNLVLAM 240

|||||

Db 181 PLFVFNAGMFYFENLSTYHNLLTETVKIVPPTLFAEQDFLNMYPKDIYKPIPPVYNLVLAM 240

|||||

QY 241 LWRHPENIELDQVVKVHYCAAGAKPWRFTEENMDREDIKMLVKKWMDIYNDESIDYKN 300

|||||

Db 241 LWRHPENIELDQVVKVHYCAAGAKPWRFTEENMDREDIKMLVKKWMDIYNDESIDYKN 300

|||||

QY 301 VVIGDSHKKOOTLQOFTLEALSEAGALQYVKAPSAA 335



Db 178 PLYFAGMEVHEPGLGTAKDLALVVTPTTFAEQDFLNMFFRQYKPIPNVNLVLAM 237  
QY 241 LMRHPENIELDQVKKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWDIYNDESIDYKN 300  
Db 238 LMRHPENVDLDQVKKVHYCAAGSKPWRFTGKEENMNRNEDIKMLVKRWMDIYNDESIDYKE 297  
QY 301 VVIGSHKKQOQLQOIFIEALSEAGALQYVKAPSA 335  
Db 298 ----EEDNADAEASQPMRTALAEAGAVKYPAPSA 328  
RESULT 5  
H96629  
hypothetical protein F8A5.2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H96629  
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: H96629  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-334 <STO>  
A:Cross-references: GB:AE005173; NID:g2462751; PIDN:AAB71970.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F8A5.2  
A:Map position: 1

Query Match 71.5%; Score 1298.5; DB 2; Length 334;  
Best Local Similarity 70.7%; Pred. No. 3.9e-100;  
Matches 241; Conservative 35; Mismatches 52; Indels 13; Gaps 4;  
QY 1 MAPEINTK---LTVPVHSAATGCKRAYVTFLAGTDYVGVVGLAKGLRKAASKYPLVVA 57  
Db 1 MAPEISVPMYLSERAHQAP--PRAYVTFLAGNDYVGVVGLAKGLRKYKSAIPLVVA 58  
QY 58 VLPDPVEDHRKOLVDQGVVKEIEPVYPENOTEFAMAYVINYVSKLRWFEVYKNMIY 117  
Db 59 MLPDPVEEHLRLSGQCVREIEPVYPPDQVEFAMAYVINYVSKLRWFEVYKNMIY 118  
QY 118 LDGDIQVFNIDHLPNGQFYAVMDCFCETWNSHSPQYKIGYCCQCPDKVTPWPEAKLG 177  
Db 119 LDADIQVFNIDHLPDLSDAYVAVMDCFCETWNSHSPQYKIGYCCQCPDKVTPWPEAKLG 178  
QY 178 PRPPLFYFAGMEVYEPNLSYHNLETVKIVPPTLFAEQDFLNMFFKDIYKPIPPVNLV 237  
Db 179 PRPPLFYFAGMEVYEPNLSYHNLETVKIVPPTLFAEQDFLNMFFKDIYKPIPPVNLV 238  
QY 238 LAMLRHPENIELDQVKKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWDIYNDESID 297  
Db 239 LAMLRHPENVELEKVVHYCAAGSKPWRFTGEEENMDREDIKMLVKRWMDIYNDESID 298  
QY 298 YKNVIGDSHK---KQOTLQOIFIEALSEAGALQYVKAPSA 335  
Db 299 FKSKIPADAEETVTKSSILASVLEP-----EMTYPPAPSA 334

RESULT 6  
T06009  
hypothetical protein T25K17.60 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999

C:Accession: T06009  
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15382  
A:Accession: T06009  
A:Molecule type: DNA  
A:Residues: 1-361 <BEV>  
A:Cross-references: EMBL:AL049171; GSPDB:GN00062; ATSP:T25K17.60  
A:Experimental source: cultivar Columbia; BAC clone T25K17  
C:Genetics:  
A:Gene: ATSP:T25K17.60  
A:Map position: 4  
A:Introns: 136/3; 244/3; 289/3  
Query Match 56.0%; Score 1198.5; DB 2; Length 361;  
Best Local Similarity 65.0%; Pred. No. 8.6e-92;  
Matches 221; Conservative 31; Mismatches 61; Indels 27; Gaps 2;  
QY 21 KRAYVTFLAGTDYVGVVGLAKGLRKAASKYPLVVAVLPDPVEDHRKOLVDQGVVKEI 80  
Db 24 KRAYVTFLAGNDYVGVVGLAKGLRKYKSAIPLVVAIPLDPVEEHLRLSGQCVREI 83  
QY 81 EPVYPENOTEFAMAYVINYVSKLRWFEVYKNMIY 115  
Db 84 EPVYPENOTEFAMAYVINYVSKLRWFEVYKNMIY 115  
QY 116 IYLDGDIQVFNIDHLPNGQFYAVMDCFCETWNSHSPQYKIGYCCQCPDKVTPWPEAK 175  
Db 144 IYLDGDIQVFNIDHLPNGQFYAVMDCFCETWNSHSPQYKIGYCCQCPDKVTPWPEAK 203  
QY 176 LGPKPPLFYFAGMEVYEPNLSYHNLETVKIVPPTLFAEQDFLNMFFKDIYKPIPPVNLV 235  
Db 204 LGSPPPVYFAGMEVYEPNLSYHNLETVKIVPPTLFAEQDFLNMFFKDIYKPIPPVNLV 263  
QY 236 LVLAMLRHPENIELDQVKKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWDIYNDES 295  
Db 264 LVAMLRHPENIELDQVKKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWDIYNDES 323  
QY 296 LDYKNVIGDSHK---KQOTLQOIFIEALSEAGALQYVKAPSA 335  
Db 324 LDYKNVIGDSHK---KQOTLQOIFIEALSEAGALQYVKAPSA 361

RESULT 7  
T02295  
hypothetical protein T13D8.32 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 22-Oct-1999  
C:Accession: T02295  
R:Vyotskaja, V.S.; Schwartz, J.R.; Kwan, A.; Toriumi, M.; Yu, G.; Oji, O.; Liu, S.;  
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel  
submitted to the EMBL data library, June 1998  
A:Description: Arabidopsis thaliana chromosome 1 BAC T13D8 sequence.  
A:Reference number: Z14649  
A:Accession: T02295  
A>Status: translated from GB/EMBL/DDBT  
A:Molecule type: DNA  
A:Residues: 1-345 <VYS>  
A:Cross-references: EMBL:AC004473; NID:g3108025; PID:g3249091; GSPDB:GN00059; ATSP:T1  
C:Genetics:  
A:Gene: ATSP:T13D8.32  
A:Map position: 1  
A:Introns: 119/3; 227/3; 272/3  
Query Match 61.1%; Score 1110; DB 2; Length 345;  
Best Local Similarity 59.9%; Pred. No. 1.8e-84;  
Matches 203; Conservative 46; Mismatches 64; Indels 26; Gaps 3;  
QY 20 EKRAYVTFLAGTDYVGVVGLAKGLRKAASKYPLVVAVLPDPVEDHRKOLVDQGVVKEI 79  
Db 10 KERAYVTFLAGNDYVGVVGLAKGLRKYKSAIPLVVAIPLDPVEEHLRLSGQCVREI 69

QY 80 IEPYCPENOTEFAMAYVYVINYNSKLRW-----EFVEYNKMIYL 118  
 ||||:|:| :| |||:||||||| | |||||  
 Db 70 IEPVHPDSDAYARAYIINYNSKLRITWNSVYIYRLHENESLRLSLNNFEYNKMIYL 129  
 ||||| ||| |||:| :| :| ||||| |||:| |||||  
 QY 119 DGIQVFDNIDHLDLPNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTPWPAKLG 178  
 | ||||| ||| |||:| :| :| ||||| |||:| |||||  
 Db 130 DADIQVFGNIDDLFDQDQGLHGVLSFCFCKINSYPLISIGYCYQCEKVVWPAEMESA 169  
 ||||| ||| |||:| :| :| ||||| |||:| |||||  
 QY 179 KPLPYFNAGMYVEPNLSVTHNLETVKVIYVPTLFAEQDFLNMYFKDIYPIPPVYNVL 238  
 | ||||| ||| ||| :| |||:| :| ||||| ||| :| |||:| |||||  
 Db 190 PPSPYFNAGMFVEPNPLTVESLLQTLQVPTPTFAEQDFLNMEFFGKVKFVPSVYNLIL 249  
 ||||| ||| |||:| :| :| ||||| |||:| |||||  
 QY 239 AMLWRHPENITELQVKKVHVHCAAGAPWRFTGEEENMDREDIKMLVKKWDIYNDESLDY 298  
 :| ||||| :| ||| :| ||||| ||| :| ||||| |||:| |||||  
 Db 250 SVLWRHPGKVDLESVKVHVHCVCPGSKPWRITGEPNMDREDVKMLIKKWDIYNDES LDF 309  
 ||||| ||| |||:| :| :| ||||| |||:| |||||  
 QY 299 KNVVIGDGHKKQOTLOQ--FIEALSEAGALQYVKAPSA 335  
 | :| :| :| |||||  
 Db 310 KPKSPAD--LEATVLESTIIASVT EAPLSYSPAASAA 345  
 |||||

RESULT 8  
 T00444  
 hypothetical protein T14N5.1 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 22-Oct-1999  
 C:Accession: T00444  
 R:Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo,  
 J.; Vysotskaia, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.  
 submitted to the EMBL data Library, September 1998  
 A:Reference number: Z14152  
 A:Accession: T00444  
 A:Status: translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1201 <FED>  
 A:Cross-references: EMBL:AC004260; NID:g3176694; PID:g3540195; GSPDB:GN00059; ATSP:T14N5.1  
 A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Gene: ATSP:T14N5.1  
 A:Map position: 1  
 A:Introns: 251/1; 569/1; 601/2; 858/1

RESULT 9  
JC4695  
glycogenin glucosyltransferase (EC 2.4.1.186) - human  
N:Alternate names: glycogenin  
N:Contains: glycogen(starch) synthase, glycogenin subunit  
C:Species: Homo sapiens (man)  
C:Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 11-Jun-1999  
A:Accession: JC4695; G01948; S45141  
R:Barbetti, F.; Rocchi, M.; Bossolasco, M.; Cordera, R.; Sbraccia, P.; Finelli, P.; C  
Biochem. Biophys. Res. Commun. 220, 72-77, 1996  
A:Title: The human skeletal glycogenin gene: cDNA, tissue expression, and chro  
A:Reference number: JC4695; MUID:96176958  
A:Accession: JC4695  
A:Molecule type: mRNA  
A:Residues: 1-333 <BAR>  
A:Cross-references: EMBL:U44131; NID:g1174166; PIDN:AAB00114.1; PID:g1174167  
A:Experimental source: skeletal muscle  
R:Rodriguez, I.R.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: G08914  
A:Accession: G01948  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-333 <POD>  
A:Cross-references: EMBL:U31525; NID:g976399; PIDN:AAB09752.1; PID:g976400  
R:Jeffers, H.; Wiemann, S.; Ansorge, W.  
submitted to the EMBL Data Library, June 1994  
A:Description: Cloning and sequencing of a cDNA encoding human glycogenin.  
A:Reference number: S45140  
A:Accession: S45141  
A:Molecule type: mRNA  
A:Residues: 1-203, 'KMSQEPHYICPLGRSOLWSHSLYPRKNCR', 'NDGNRALLIWEQIPLTTSRGNLTSSR', 'NT  
A:Cross-references: EMBL:X79537; NID:g456894; PIDN:CAA56073.1; PID:g496895  
C:Genetics:  
A:Gene: GDB:GYG  
A:Cross-references: GDB:1326953  
A:Map position: 3q24-3q25.1  
C:Function:  
A:Description: catalyzes the alpha-glucosylation of itself on a specific tyrosine res  
osyl-glycogenin by UDPglucose to produce glycogen and UDP  
A:Pathway: glycogen/starch biosynthesis  
A:Note: required to initiate the synthesis of glycogen  
C:Superfamily: glycogenin  
C:Keywords: acetylated amino end; glycogen/starch biosynthesis; glycoprotein; glycosy  
F:2/Modified site: acetylated amino end (Thr) (in mature form) #status predicted  
F:195/Binding site: carbohydrate (Tyr) (covalent) #status predicted

Query Match	13.3%	Score	241;	DB	1;	Length	333;
Best Local Similarity	24.8%;	Pred. No.	2.4e-12;				
Matches	78;	Conservative	43;	Mismatches	120;	Indels	74;
Gaps	11;						
QY	22	RAYVFLAGTCDYKGVVGLAKGLRKAKSKYPIVAVVLDPVPEDHRKQL--VDGCVVKE	79				
		: : :	: : :	: :	: :	: :	: :
Db	4	QAFVT-LTTNDAYAKGALVGLGSSLKQHRTTTRLVVLATPQVSDSMRKVLETFVDEVIMVD	62				
QY	80	IEPVYPENOTEFAMAYVINYSLRIWEFYENKMTYLDGDIQVFDNIHLFDLPNGQF	139				
		: :	: :	: :	: :	: :	: :
Db	63	VLDGSDSAHLFLMKRPELGVTLTLKHCWSLTQYKCVFMDADTVLANIDDLFDEE---	119				
QY	140	YAVMDCFEKTSWHSPOYKIGYCOQCDKVPWPEAKLGPKPLFYFNAGMFVYEPNLSYH	199				
		: :	: :	: :	: :	: :	: :
Db	120	-----LSAAPDP-CWPPDC-----FNSGVFYQPSVETYN	147				
QY	200	NLETVKIVPTLFAEQDFLNNYFK-----DIYKPIPPVYNVLVLAHLWRHPENIEL--DQ	252				
		: :	: :	: :	: :	: :	: :
Db	148	QLLHLASEQGSFGDGQGIILTFESSWATTDIRKHLFFIYNLSSISYSLYPAFKVFGAS	207				
QY	253	VKVYHYCAAGAKWRFT-----GREENMDREDIKMLVKKKWWDIYNDESILDYKNV	301				
		: : :	: :	: :	: :	: :	: :
Db	208	AKVYHFLGR-VKPNWNYTDPKTKSVKSEADPNMTHPEFLI---WWNIFTNVL-----	258				
QY	302	VIGDSHKKQOTLOOF	316				



RESULT 12  
T40489  
hypothetical protein SPBC4C3.08 - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: *Schizosaccharomyces pombe*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T40489  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.  
submitted to the EMBL data Library, February 1998  
A:Reference number: Z21910  
A:Accession: T40489

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-372 <WOO>

A:Cross-references: EMBL:AL021730; PIDN:CAAL6830.1; GSPDB:GN00067; SPDB:SPBC4C3.08

A:Experimental source: strain 972h; cosmid c4C3

C:Genetics:

A:Gene: SPDB:SPBC4C3.08

A:Map position: 2

Query Match 12.4%; Score 225.5; DB 2; Length 372;

Best Local Similarity 25.3%; Pred. No. 5.4e-11;

Matches 89; Conservative 44; Mismatches 128; Indels 91; Gaps 18;

QY 2 APEINTKLVVPHSATG--GKRAVVTFL---AGTGD-----YKGVVGLAKGL 45

DB 38 SPEIQRS---VYTLGLAPSKMAFVMTLVRAANGENEVENTQODWYNSRLLVRL 93

QY 46 RK---AKSKYPLVAVLPDPEDHRKQLVDQCVVKEIEPVVPE-----NOTEFAMAYY 97

DB 94 VKFPKTSKYPVVVLAAMKIDQWKLQDQEDGAIKVVVDPLYAHEVVDVNDIALLSRW 153

QY 98 VVNSKLRITWEFVYKMYILGDDIQVFNIDHFLDL-----PNCQFYAV-MD 144

DB 154 SMFTKLRVFEYEDRICFLDSILPIKMKDKVDFVHQLSYKSDSVLFPPTFLFYKPRRS 213

QY 145 CFCEKTSWSPQYKIGYCOQCP-----DKVTWPEAKLGPKPPL--YFNAGMFVYEPNL 195

DB 214 IFWRRTEFEFAYGLTRDDLVPYVFAVSDPGMWHET---PPPKDYFNAGLFVFKPLK 269

QY 196 STVHNLETVKIVPPPLF-----AEQDFLMYFKDIYKPIPPVYNLVLAHLWRH----- 244

DB 270 AHYKRLMALARE--PKLYDNANMEOSSLNF-----AYNSAGAFWPESLDWTFN 316

QY 245 ---PENIELDQVYVHYCAAGAKPRFTEGEENMDREDIKMLVKKWMDIYND 293

DB 317 GLWARKNDLPYLRKVH-----GKHQWPEG-SLGYDEDYSKL-----WDAFQE 358

RESULT 13

A45729

glycoengin homolog, sqdb 3'-region - Rhodobacter sphaeroides

C:Species: Rhodobacter sphaeroides

C>Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 08-Oct-1999

C:Accession: A45729; S27663

J. Benning, C.; Somerville, C.R.

J. Bacteriol. 174, 6479-6487, 1992

A:Title: Identification of an operon involved in sulfolipid biosynthesis in Rhodobacter

A:Reference number: A45729; MUID:93015699

A:Accession: A45729

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-260 <BEN>

A:Cross-references: EMBL:M89780; NID:g152038; PIDN:AAA73224.1; PID:g152040

C:Genetics:

A:Start codon: GTG

Query Match

Best Local Similarity 11.3%; Score 205.5; DB 2; Length 260;

Matches 66; Conservative 46; Mismatches 93; Indels 79; Gaps 12;

QY 33 DYKGVVYGLAKGLRKAQSKYPLVAVLPDPEDHRKQLVDQCVVKEIE--PVYPPENQT 90

DB 5 DYALGARALLRSLSGTTADRVV-LHTDVPPEALAPRALGARLVVRVLELTPSPFNAA 63

QY 91 EFAMAYY-----VINSKLRITWEFVYKMYILGDDIQVFNIDHFLDL 134

DB 64 HAREALHARAFAFTKGGKPPFHTPLDNFAKLRLWQLVDYRSVFIDADALVLRNDRLEFDY 123

QY 135 PNCQFYAVMDCFEKTSWSPQYKIGYCOQCPDKVTWPEAKLGPKPPLY-----FNA 186

DB 124 PE-----FC-----AAPNVYESLSDFHRMS 144

QY 187 GMFVYEPNLSTYHNLETVKIVPPTLF---AEQDFLMYFKDIYKPIPPVYNLVLAHLWR 243

DB 145 GVFTARESTDYARMLEALDV--PGAEMRRTDQSFLOQFFPD-WQGLPVCNMLQYVWFA 201

QY 244 HPENIELDQVYVHYCAAGAKPRFTEGEENMDREDIKMLVKKW 287

DB 202 MPGLWSWEQIRILHF--QYKPMQ---AHDKADR--LRPLIDLW 238

RESULT 14

T05984

hypothetical protein F17M5.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999

C:Accession: T05984

R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15263

A:Accession: T05984

A:Molecule type: DNA

A:Residues: 1-333 <BEV>

A:Cross-references: EMBL:AL035678; GSPDB:GN00062; ATSP:F17M5.90

A:Experimental source: cultivar Columbia; BAC clone F17M5

C:Genetics:

A:Gene: ATSP:F17M5.90

A:Map position: 4

A:Introns: 193/2

Query Match

Best Local Similarity 11.3%; Score 205; DB 2; Length 333;

Matches 70; Conservative 50; Mismatches 110; Indels 66; Gaps 11;

QY 20 EKRAYVTFLAGTDYKGVVGLAKGLRKAQSKYPLVAVLPDPEDHRKQLVDQCVVKE 79

DB 36 KREAYVTVLHSSSYVCGATLAQSLLOTWTKRDLILLHDDSSITKRLAALAAAGKLR 95

QY 80 IEPVYPPENOTEFAMAYYVINYKLRITWEFVYKMYILGDDIQVFNIDHFLDLPLNGQF 139

DB 96 IIRIRNPLAEKD---SYNEYNYSKFRMLQTLTDYDKVIFADIDIVLRNLDLLEFFHQ--- 149

QY 140 YAVMDCFEKTSWSPQYKIGYCOQCPDKVTWPEAKLGPKPPLYFNAGMFVYEPNLSTYH 199

DB 150 ---MSATGNDVW-----IYNSGIMVIEPSNCTFT 175

QY 200 NLETVKIVPPTLFAEQDFLMYFKDIYKPIPPVYNLVLAHLWR-----RHPEN---IE 249

DB 176 TIMSORSEIYSYNGDQGLYLNELFVWHR-LPRVNF-LKNFWSNTTKERNKNNLFAAE 233

QY 250 LDQVYVHYCAAGAKPW-----RFTGEENMDREDIKMLVKKWMDIYNDESLD 297

DB 234 PPQYVAVHY--LGKWPWCYRDYDCNYDVEQLVYASDAAHV--RWWKVH--DSMD 283

RESULT 15

T31001

hypothetical protein F56B6.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T31001

R:Stellies, L.; Stellye, L.

submitted to the EMBL Data Library, September 1999

A:Description: The sequence of C. elegans cosmid F56B6.

A:Reference number: Z20957

A:Accession: T31001

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-429 <STEL>

A:Cross-references: EMBL:U64599; PIDN:AAB04562.1

A:Experimental source: strain Bristol N2

C:Genetics:

A:Map position: X

A: Introns: 41/3; 105/3; 146/1; 202/2; 245/3; 371/3; 390/3  
A: Note: F56B6.4

```
Query Match      10.8%; Score 197; DB 2; Length 429;
Best Local Similarity 23.6%; Pred. No. 1.5e-08;
Matches 75; Conservative 46; Mismatches 101; Indels 96; Gaps 16;

QY 23 AVVTFLAGTGYVKGVGGLAKLRKAKSKYPLVAVLPDPEDHRKQLVDQGVVKEIEP 82
   ||| || | : | | : | | | : : | | | | : | : |
Db 4 AWIT-LATNDNYAOGALVLVHSLRTAGTRKIHCLISNEVSAPVRKKQLEEHFDDVSIVD- 61
   | : : : : : : : : : : : : : : : : : : : : : : : :

QY 83 VYPPEOTEFAMAY---YVINYSKLRIWEFVEYNKMIYLDGDIQVFDNIDHFLDLPNGQF 139
   | : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 VNSNDSNLRLLIERPDLGVFTKLHCWRLTQYTKCVFLDADTLVLRNADELFTRPD--F 119
   | : : : : : : : : : : : : : : : : : : : : : : : :

QY 140 YAVMDFCFCEKTSWHSHPQYKIGYCCQCPDKVTWPEAKLGPKPPLYFNAGMFVYEPNLSYH 199
   | | : : : : : : : : : : : : : : : : : : : : : : :
Db 120 SAASD-----IGWPDG-----FNSGVFYVPNNETIR 146
   | : : : : : : : : : : : : : : : : : : : : : : : :

QY 200 NLETVKIVPPTLEA-----EODELNMYF---KDIYKP--IPPVYNLVLA MLW--- 242
   ||| || | : | | : | | | : | | : | | : | : | :
Db 147 QLVD-----FAVTHGSYDGGQGLNDFFSNWRDLPSHRLPFIYNTAGAFYYA 197
   | : : : : : : : : : : : : : : : : : : : : : : : :

QY 243 ----RHPENIELDQVKVHYCAAGAKPWR-----FTGEEENMDREDIKMLVKKWDIY-- 291
   | : | : : : : : : : : : : : : : : : : : : : : : :
Db 198 RAYKRYGAN----TKIVHFIGS-VKPHGSAAVHTGEH-----FQWQKIYHA 240
   | : | : : : : : : : : : : : : : : : : : : : : : :

QY 292 --NDESLDYKNVVICGDSH 307
   | | : : : : : : : : : : : : : : : : : : : : : :
Db 241 HVNHTSRTNEHAAVFPESH 258
   | | : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: July 2, 2002, 12:09:34  
Job time: 181 sec





Db 2 MGPVNSE---KKAIAAAKRRAYVTFLAGDGYWGVVGLAKGLRRVRSAYPLVVAVLP 57  
QY 61 DVPEDHRKQLVDGCVKETEYPENQTEFAMAYVINYSLRIWEFVEYNKMIYLDG 120  
Db 58 DVPCEHRKQLVGGCVREIQVYYPESQTFAMAYVINYSLRIWEFVEYERMYLDA 117  
QY 121 DIQVFDNIDHLFDPNGQFVAVMDCFCCKTWSHSPQKIGYCCQCPDKVTPWPEAKLGPKP 180  
Db 118 DIQVFDNIDHLFDPNGQFVAVMDCFCCKTWSHSPQKIGYCCQCPDKVTPWPEAKLGPKP 177  
QY 181 PLYENAGMFVEYNLSYHNLLTVPPTLFAEODELNMVFKDIYKPTPPVYNLVLAM 240  
Db 178 PLYENAGMFVEYNLSYHNLLTVPPTLFAEODELNMVFKDIYKPTPPVYNLVLAM 237  
QY 241 LWRHPENIDLVKVVHYCAAGAKPWRFTGEEENMDREDIKMLVKWWDIYNDESIDYKN 300  
Db 238 LWRHPENIDLVKVVHYCAAGAKPWRFTGEEENMDREDIKMLVKWWDIYNDESIDYKN 297  
QY 301 VVIGDSHKKOOTTQOFTFALSEAGALQYVKAPSA 335  
Db 298 ---EEDNADEASQPMRTALAEAGAVKYPAPSA 328

## RESULT 2

US-09-073-297-2  
; Sequence 2, Application US/09073297  
; Patent No. 6255114  
; GENERAL INFORMATION:  
; APPLICANT: LIGHTNER, JONATHAN EDWARD  
; TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: U.S.A.  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.50 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD VERSION 7.0A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/073,297  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/852615  
; FILING DATE: MAY 7, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MAJARIAN, WILLIAM R.  
; REGISTRATION NUMBER: 41,173  
; REFERENCE/DOCKET NUMBER: BB-1083-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-992-4926  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 346 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-073-297-2

Query Match 14.6%; Score 265.5; DB 4; Length 346;  
Best Local Similarity 27.6%; Pred. No. 3.6e-21;  
Matches 81; Conservative 45; Mismatches 98; Indels 69; Gaps 11;

QY 21 KRAYVTFLAGTGDYKVGWGLAKGLRKAQSKYPLVAVLPDVPEDHRKQLVDGCVVKEI 80  
Db 7 REATATILHSASEVCAITAAQSIQAGSTRDLVILVDDTISDHRKGLGESAGWVRIL 66

QY 81 EPPVPPENQTEFAMAYVINYSLRIWEFVEYNKMIYLDGDIQVFDNIDHLFDPNGQFY 140  
Db 67 QRIRNPKEARD---AYNEWNTSKFRLWQTDYKVFIDADLLIIRNIDFLFALP----- 118  
QY 141 AVMDCFCEKTSWHSPOYKIGYCCQCPDKVTPWPEAKLGPKPLPYENAGMFVEYNLSYHN 200  
Db 119 -----EIT-----ATGNATL-----FNSGVWVIEPSNCTFRL 146  
QY 201 LLETVKIYVPTLFAEQDFLNMVFKDIYKPTPPVYNLVAMLRHPENIDELDOVK----- 254  
Db 147 LMEHDEITSYNGDGYLNEIFTWHR-IPKHMNF-LKHEWEGE-----EEVRAKKTRL 200  
QY 255 -----VVHYCAAGAKPWR-FTGEEENMDREDIKMLVK-----KWWDIYN 292  
Db 201 FGANPPVLVLYHY--LGRKPWLCEFRDYDCNNWNVETLREFASDVAHARWVKVHN 251

## RESULT 3

US-09-073-297-3  
; Sequence 3, Application US/09073297  
; Patent No. 6255114  
; GENERAL INFORMATION:  
; APPLICANT: LIGHTNER, JONATHAN EDWARD  
; TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: U.S.A.  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.50 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD VERSION 7.0A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/073,297  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/852615  
; FILING DATE: MAY 7, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MAJARIAN, WILLIAM R.  
; REGISTRATION NUMBER: 41,173  
; REFERENCE/DOCKET NUMBER: BB-1083-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-992-4926  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 333 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-073-297-3

Query Match 13.3%; Score 241; DB 4; Length 333;  
Best Local Similarity 24.8%; Pred. No. 1.9e-18;  
Matches 78; Conservative 43; Mismatches 120; Indels 74; Gaps 11;

QY 22 RAYVTFLAGTGDYKVGWGLAKGLRKAQSKYPLVAVLPDVPEDHRKOL--VDGCVVKE 79  
Db 4 QAFVT-LTTNDAYAGALVGLSSLAQHRTTRRLVLAATPQVSDSNRKVLETFVEVINWD 62  
QY 80 IEPVPPENQTEFAMAYVINYSLRIWEFVEYNKMIYLDGDIQVFDNIDHLFDPNGQF 139  
Db 63 VLDSGSAHLTLMKRPELGVTLTKLHCNSLTQYKCVFMDADTLVLANIDDLDFREE--- 119

QY 140 YAVMDCFEKTSWHSPOYKIGYCOCPDKVTWPEAKLGPDPPLYNAGMFVPEPNLSTYH 199  
Db 120 -----LSAAPDP-GWPCD-----FNSGVFYQPSVETYN 147  
QY 200 NLETVKIVPPTLFAEQDFLNMFK-----DIYKPIPPVYNVLVLAHLRHPENIEL--DQ 252  
Db 148 QLLHLASEQSGFDGQDQILNTFFSSWATTDIRKHLPIYLLSSISYSLPAFKVFGAS 207  
QY 253 KVVVHYCAAGAKPWFET-----GEBENMDREDIKMLVKKWMDIYNDESIDYKNV 301  
Db 208 AKVHFHGLR-VKPMNYTVDPKTSVKSEADPNMTHPEFLIL---WNFIETNVL----- 258  
QY 302 VIGDSHKKQOTLQQF 316  
Db 259 -----PLLQQF 264

RESULT 4  
US-09-073-297-5  
; Sequence 5, Application US/09073297  
; Patent No. 6255114  
; GENERAL INFORMATION:  
; APPLICANT: LIGHTNER, JONATHAN EDWARD  
; TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: U.S.A.  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.50 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD VERSION 7.0A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/073,297  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/852615  
; FILING DATE: MAY 7, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MAJARIAN, WILLIAM R.  
; REGISTRATION NUMBER: 41,173  
; REFERENCE/DOCKET NUMBER: BB-1083-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-992-4926  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 566 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-09-073-297-5

Query Match 12.5%; Score 227; DB 4; Length 566;  
Best Local Similarity 22.3%; Pred. No. 1.5e-16;  
Matches 69; Conservative 53; Mismatches 91; Indels 96; Gaps 10;  
QY 15 SATGGEKRAYTFLAGTDGYVGVVGLAKGLRKAQKSYPLVAVLPDVPDHRKQLVDQ 74  
Db 271 SSFTAQRVAYVTLHSSEVYVCGATLAQSIQSGSTKDMILLHDDSDITNLSLGLSLAG 330  
QY 75 CVVKEIEPVYPENQTERFAMAYVINYSKLRWFEVYKNMYLDGDIQVFNIDHLPDL 134  
Db 331 WKLRRVERIRSPFSKKR---SYNEWNYSKLRVQVTDYDKLVFIDADFIIVKNIDYLF-- 385

QY 135 PNGOFYAVMDCFEKTSWHSPOYKIGYCOCPDKVTWPEAKLGPDPPLYNAGMFVPEPN 194  
Db 386 -----SYPLSAAGNNKVLFNNSGVVLEPS 410  
QY 195 LSTYHNL-LBTVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNVLVLAHLRHPENIEL--DQ 244  
Db 411 ACLFEDLMKLSFKTSYN-GGDOGFLNEYF-----VMMHRLSKRLNTM 452  
QY 245 -----PENIELDQVKKVVHYCAAGAKPWR-FTGEBENMDREDIKMLVK--- 285  
Db 453 KYFGDESRHDKARNLPELNE-----GIHY--LGLKPMRCYRDYDCNMDLKRTRVYASESV 505  
QY 286 --KWMDIYN 292  
Db 506 HARWKVYD 514

RESULT 5  
US-09-073-297-14  
; Sequence 14, Application US/09073297  
; Patent No. 6255114  
; GENERAL INFORMATION:  
; APPLICANT: LIGHTNER, JONATHAN EDWARD  
; TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: U.S.A.  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.50 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD VERSION 7.0A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/073,297  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/852615  
; FILING DATE: MAY 7, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MAJARIAN, WILLIAM R.  
; REGISTRATION NUMBER: 41,173  
; REFERENCE/DOCKET NUMBER: BB-1083-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-992-4926  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 71 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; CLONE: rl0n.pk0027.fl1  
; US-09-073-297-14

Query Match 12.4%; Score 225; DB 4; Length 71;  
Best Local Similarity 59.5%; Pred. No. 1e-17;  
Matches 44; Conservative 13; Mismatches 13; Indels 4; Gaps 1;  
QY 1 MAPEINTKLIVPHSATGGEKRAYTFLAGTDGYVGVVGLAKGLRKAQKSYPLVAVLP 60  
Db 2 MGPNVSE-----KKALAAAKRRAYTFLAGDGYWKGVLAKGLRVRVSAYPLVAVLP 57  
QY 61 DVPEDHRKQLVDQ 74

[illegible]

```

12-001
US-073-297-16
App-Location US/09073297
Sequence 16, Application 16
Patent No. 6255114
GENERAL INFORMATION:
APPLICANT: LIGHTNER, JONATHAN EDWARD
TITLE OF INVENTION: STARCH BIOSYNTHESIS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESS: E. I. DU PONT DE NEMOURS & CO.
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE

```



OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD VERSION 7.0A  
CURRENT APPLICATION DATA: US/09/073.297  
APPLICATION NUMBER: 08/852615  
FILING DATE: MAY 7, 1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
ATTORNEY/AGENT INFORMATION:  
NAME: MAJARIAN, WILLIAM R.  
REGISTRATION NUMBER: 41,173  
REFERENCE/DOCKET NUMBER: BB-1083-B  
TELEPHONE: 302-992-4926  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 132 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
CLONE: rrl.pk0070.e9  
US-09-073-297-16

Query Match 8.9%; Score 161; DB 4; Length 132;  
Best Local Similarity 27.2%; Pred. No. 3.7e-10;  
Matches 43; Conservative 27; Mismatches 48; Indels 40; Gaps 4;  
Qy 66 HRKQLVDGCVVKEPYYPENQTEFAMAYVINYSLRIWEFVEYKMIYLDGDIQVF 125  
Db 1 HRKLEAGAGWVYRQIRNPKAERD---AYNEWNYSKFRLWQLTDYDKILFIDADULLIL 57  
Qy 126 DNIDHFDLPNGQFVAVMDCFCCEKTSHPQYKIGYCCQCPKVTWPEAKLGPKPPPLYFN 185  
Db 58 RNVDFLFAMP-----EIT-----ATGNNAFL-----FN 80  
Qy 186 AGMFVYEPNLSYHNLLETVKIVPTLFAEQDFLNNYF 223  
Db 81 SGVMVIEPNSCTFQLLMHDHNEITSYNGDQGYLNEIF 118

RESULT 9  
US-09-073-297-12  
Sequence 12, Application US/09073297  
Patent No. 6255114  
GENERAL INFORMATION:  
APPLICANT: LIGHTNER, JONATHAN EDWARD  
TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: U.S.A.  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD VERSION 7.0A  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073.297  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/852615  
FILING DATE: MAY 7, 1997  
ATTORNEY/AGENT INFORMATION:

NAME: MAJARIAN, WILLIAM R.  
REGISTRATION NUMBER: 41,173  
REFERENCE/DOCKET NUMBER: BB-1083-B  
TELEPHONE: 302-992-4926  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
CLONE: ctaln.pk0013.e6  
US-09-073-297-12

Query Match 8.0%; Score 145.5; DB 4; Length 93;  
Best Local Similarity 34.0%; Pred. No. 1.2e-08;  
Matches 32; Conservative 19; Mismatches 40; Indels 3; Gaps 1;  
Qy 21 KRAYVFLAGTDYVKGVLGAKLGRKAKSKYPLVAVLDPVDPEDHRKQLVDGCVVKEI 80  
Db 3 REAYATILHSASEYVCGAITAQAQSIQAQSTRLVILVDDTISDHRKGLSAGWKVRIL 62  
Qy 81 EPVYPPENQTEFAMAYVINYSLRIWEFVEYK 114  
Db 63 QRIRNPKAERD---AYNEWNYSKFRLWQLTDYDK 93

RESULT 10  
US-08-312-387B-4  
Sequence 4, Application US/08312387B  
Patent No. 5545553  
GENERAL INFORMATION:  
APPLICANT: Gotschlich, Emil C.  
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF  
OLIGOSACCHARIDES, AND GENES ENCODING THEM  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/312,387B  
FILING DATE: July 7, 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-095  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 306 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-312-387B-4

```

Db      80  ITTYARLKLGEYIADCDKVLVLDFTDVLVRDGLKPLMDTDLGGNMGACIDLFVER----- 134
Qy      155  PQYKIGYCCQCPDKVTWPFAKLG-PKPPLYFNAGMFVYEPNLSYHNL-----LETVK 206
Db      135  --QEGYKQ-----KIGMADEGYFNAGVLLINLKKWRRHDFPKMSCWEVQYK 180
Qy      207  IVPPTLFAEQDFLNMYK-----DIYKIPPPYINVL-AMLWRHPENIELDQ----- 252
Db      181  DV--MQYDQDILNGLFKGGVCYANSRNFEMPTNYAFMANGFASRHTDPLDRTNTAMP 238
Qy      253  VKVHYCAAGAKPW-----RFT---GEENMDRE-----DIKMLVKW 287
Db      239  VAVSHYCGS-AKPWHRCTVWGAERFTELAGSLTTVPENRGKLAVPPTKCMLOQR 293

RESULT 12
US-08-683-458-4
; Sequence 4, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-458-4

Query Match 7.4%; Score 134; DB 1; Length 306;
Best Local Similarity 25.0%; Pred. No. 1.4e-06;
Matches 59; Conservative 38; Mismatches 71; Indels 68; Gaps

Qy      98  VINYSKLRIWEFY-EYNKMIYLDGDIQVFDNIDHFLDPLNGOFY--AVMDCFCEKTWHS 154
Db      80  ITTYARLKLGEYIADCDKVLVLDFTDVLVRDGLKPLMDTDLGGNMGACIDLFVER----- 134
Qy      155  PQYKIGYCCQCPDKVTWPFAKLG-PKPPLYFNAGMFVYEPNLSYHNL-----LETVK 206
Db      135  --QEGYKQ-----KIGMADEGYFNAGVLLINLKKWRRHDFPKMSCWEVQYK 180

```

```

QY 207 IVPPTLFAEQDFLNMVFK-----DIYKPIPPVYNLVL-AMLWRHPENIELDQ----- 252
| : : : : : | | : : : | | : : : | | : : : |
Db 181 DV--MQYQODILNGLFKGVCYANSRNFMPNTYAFMANGFASRHTDPLYLDRNTAMP 238

QY 253 VKVHYCAAGAKPW-----RFT---GEEENMDRE-----DIKMLVKKW 287
| : : : : | | | : | | : | | : | | : | | : |
Db 239 VAVSHYCGS-AKPWHRDCTVWGAERFTELAGSLTTPPEWRGKLAVPPTKCMQLRW 293

RESULT 13
US-08-878-360-4
; Sequence 4, Application US/08878360
; Patent No. 5945322
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,360
; FILING DATE: 18-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/683,426
; FILING DATE:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-878-360-4

Query Match 7.4%; Score 134; DB 2; Length 306;
Best Local Similarity 25.0%; Pred. No. 1.4e-06;
Matches 59; Conservative 38; Mismatches 71; Indels 68; Gaps 14;

QY 98 VINYSLKRIWEFV-EYNKMIYLDGDIQVFNIDHLPDLPNGQFY--AVMDCFCCKTWSHS 154
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 80 ITTYARLKLGEYIADCKVLYLDTDLVLRDGLKPLWDTDLGNNWVGACIDLFVER----- 134

QY 155 PQYKIGYCQCQPKVTWPEAKLG-PKPPLYFNAGMFVYEPNLSTYHNL-----LETVK 206
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 135 ---QEGYKQ-----KIGMADGEYFYNAGVLLINLKKWRRHDIKMSCEWVEQYK 180

QY 207 IVPPTLFAEQDFLNMVFK-----DIYKPIPPVYNLVL-AMLWRHPENIELDQ----- 252
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 181 DV--MQYQODILNGLFKGVCYANSRNFMPNTYAFMANGFASRHTDPLYLDRNTAMP 238

QY 253 VKVHYCAAGAKPW-----RFT---GEEENMDRE-----DIKMLVKKW 287
| : : : : | | | : | | : | | : | | : | | : |
Db 239 VAVSHYCGS-AKPWHRDCTVWGAERFTELAGSLTTPPEWRGKLAVPPTKCMQLRW 293

RESULT 14
US-08-478-140B-4
; Sequence 4, Application US/08478140B
; Patent No. 6127153
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, KARL F.
; APPLICANT: ROTH, STEPHEN
; APPLICANT: BUCZALA, STEPHANIE L.
; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
; TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,140B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-478-140B-4

Query Match 7.4%; Score 134; DB 3; Length 306;
Best Local Similarity 25.0%; Pred. No. 1.4e-06;
Matches 59; Conservative 38; Mismatches 71; Indels 68; Gaps 14;

QY 98 VINYSLKRIWEFV-EYNKMIYLDGDIQVFNIDHLPDLPNGQFY--AVMDCFCCKTWSHS 154
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 80 ITTYARLKLGEYIADCKVLYLDTDLVLRDGLKPLWDTDLGNNWVGACIDLFVER----- 134

QY 155 PQYKIGYCQCQPKVTWPEAKLG-PKPPLYFNAGMFVYEPNLSTYHNL-----LETVK 206
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 135 ---QEGYKQ-----KIGMADGEYFYNAGVLLINLKKWRRHDIKMSCEWVEQYK 180

QY 207 IVPPTLFAEQDFLNMVFK-----DIYKPIPPVYNLVL-AMLWRHPENIELDQ----- 252
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 181 DV--MQYQODILNGLFKGVCYANSRNFMPNTYAFMANGFASRHTDPLYLDRNTAMP 238

QY 253 VKVHYCAAGAKPW-----RFT---GEEENMDRE-----DIKMLVKKW 287
| : : : : | | | : | | : | | : | | : | | : |
Db 239 VAVSHYCGS-AKPWHRDCTVWGAERFTELAGSLTTPPEWRGKLAVPPTKCMQLRW 293

RESULT 15

```

Query Match	7.4%	Score 134;	DB 4;	Length 306;
Best Local Similarity	25.0%;	Pred. No. 1.4e-06;		
Matches 59; Conservative	38;	Mismatches 71;	Indels 68;	Gaps 14;

  

QY	98	VINYSLKRIWFEV-EYNKIWIYLDGDIQVENIDHLPDLPNGQFY--AVMDFCEKTWSHS	154
		: : :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	80	IITYARLKLGEIADCDKVILDTDLVLRGLPLPMDTDLGGNWWGACIDLFEVER----	134
QY	155	POYKIGYQQCPDKVTWPPEAKLG-PKPFLYAGNMFFVEBNTSYHNL-----LETWK	206
		:   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	135	---QEGRYQ-----KGMADGEYFENAGVLLINLKWRHRHDIFKMSCEWVEYQK	180
QY	207	IYPPTLFALAQDFLNMYFK-----DIYKIPPYNYNLV-AMLRRHPENIELDQ-----	252
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	181	DV--HQYQDDILLCLFGKGVCYANSRENFMPNTYAFMANGFASRTDPLYLDRTNTAMP	238
QY	253	VKVWHYCAAGAKPW-----RFT-----GEBENMDRE-----DIKMLVKKK	287
		:   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	239	VAVSHYCGS-KAPPHRDCTVGNAERTELAGSLTTVPPEWRGKLAVPPTKCMLOQR	293

Search completed: July 2, 2002, 12:08:57  
Job time: 174 sec

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 18-MAY-1999; 99US-0134941.  
PR 19-MAY-1999; 99US-0135124.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 08-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142377.  
PR 13-JUL-1999; 99US-0143342.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 18-JUL-1999; 99US-0144086.  
PR 18-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.







DE Arabidopsis thaliana protein fragment SEQ ID NO: 14421.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX EP1033405-A2.  
PN 06-SEP-2000.  
PD  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.

```
PR 30-AUG-1999; 99US-01513303.
PR 31-AUG-1999; 99US-01514338.
PR 01-SEP-1999; 99US-01519330.
PR 07-SEP-1999; 99US-01523363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-01537758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-01540339.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-01551339.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 76.5%; Score 1389.5; DB 21; Length 333;
Best Local Similarity 77.4%; Pred. No. 9.9e-141;
Matches 254; Conservative 30; Mismatches 43; Indels 1; Gaps 1;

QY 9 LTPVHSATGGERAYVTFAGTGDYVGVGLAKGLRKAASKYPLVAVLPDVPEDHRK 58
Db 6 itkpslpsvdsrdgdyvflagngdyvkvglakgrkvksayplvavilpdpveehr 65

QY 69 QLVDOGVKWEIEPVYPENOTEFAMAYVINYSKLRWEFEYKNMIVLDGDIQVFDNI 128
Db 66 ilvdqgcivrieipvpennqtfanayvinykskirkwfkfveyskmiylldgdiqvyeni 125

QY 129 DHLFDLPNGOFYAVMOCFCBKTWSHSQYKIGYCQCQDPKVTWPEAKLGKPLPYFNAGM 188
Db 126 dhlfidpdyglyavmdcefkewtshtpqkykryccqcdkqvqpkaelgeppalyfnagm 185

QY 189 FVYEPNLSTYHNLETVKIVPPLFAEQDLNMYFKDVKYKPPYVNLVLMWRHPENI 248
Db 186 flyephletyedilrtktptpfaeqdflmnyfkkiykpilvynvlamlwrhpenv 245

QY 249 ELQGVKVHVHCAAGAKPWRFTGEENNDREDIKMLVKWMDIYNDESIDYKN-VVIGDSH 307
Db 246 elgkvkvvhvcaagskpwrytgeanmeredikmlvkkwmdlyddesldykpvtvdtc 305
```

```
QY 308 KKQOTLQOFIEALSAGALQYVKAPSAA 335
Db 306 wdlvnlkpfitalteagrlnyvtapsaa 333

RESULT 4
AAG45748
ID AAG45748 standard; Protein; 333 AA.
AC AAG45748;
XX
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 57474.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
```





QY 1 MAPEINT-KLTVPVHSA--TGGKRAYVTFLLAGTDYVGVVGLAKGLRKAASKYPLVVA 57  
 DB 1 magnittvktitdaqakvatdhrayvtflagnqdyvkvglakglrkxksmyplvva 60  
 QY 58 VLDPVPEHRKQLVDOGCVVVKEIEPVPPENOTEFAMAYVINYSLRIWEFVEYNKMIY 117  
 DB 61 vlpdpqdhrlntlsgcivrelpeypypenqcfamayvinykslriwefveyskmy 120  
 QY 118 LDGDIQVFNIDHFLDPLNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTWPPEAKLG 177  
 DB 121 ldgdiqvfndihldfplndnyfayvmdcfccktwgthqyqgycqcpkhkwp-thfg 179  
 QY 178 PKPPLFNAGMFVEPNLSTYHNLLETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLY 237  
 DB 180 pkpplfnagmfvepnlstyrldlgtvgtqptsfaeqdflnmyfkdyripnvnly 239  
 QY 238 LAMLRHPENIELDQKVHVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWMDIYNDESLD 297  
 DB 240 lamlrhpenveldkvkvhycaagkpwrytgkeenmeredikmlvkkwmdiyedetld 299  
 QY 298 YKNVVGSDSHKKQOOLQOFTIEALSEAGALQYVKAPSAA 335  
 DB 300 ynppl-----nvdkftaalmevgevkfvrapsaa 328

## RESULT 7

AAW82560  
 ID AAW82560 standard; Protein; 328 AA.

AC AAW82560;

DT 09-FEB-1999 (first entry)

XX Rice glycogenin-like water stress protein.

XX Glycogenin; water stress protein; starch biosynthesis; corn; rice;  
 wheat; plant; granules; branch chain; endosperm.

XX Oryza sativa.

XX WO9850553-A1.

PN 12-NOV-1998.

PD 06-MAY-1998; 98WO-US09201.

XX 07-MAY-1997; 97US-0852615.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

PA Lightner JE;

XX WPI; 1998-610384/51.

XX Plant glycogenin and water stress proteins - used to alter plant  
 glycogenin and water stress protein expression.

PS Disclosure; Page 33-34; 57pp; English.

XX This sequence represents a rice glycogenin-like water stress protein  
 which is used in a method to isolate novel plant glycogenin proteins.  
 CC Such proteins can be used to alter plant glycogenin and water stress  
 CC protein expression. Manipulation of glycogenin expression can be used  
 CC to alter starch biosynthesis and effects the number of starch granules  
 CC in the endosperm of corn. Overexpression or reduction of expression  
 CC of genes encoding glycogenin in corn, rice and wheat could be used to  
 CC alter branch chain distribution of the starch produced by these plants.

XX Sequence 328 AA;

Query Match

72.6%; Score 1319; DB 19; Length 328;

Best Local Similarity 71.0%; Pred. No. 3.8e-133;  
 Matches 238; Conservative 41; Mismatches 48; Indels 8; Gaps 2;  
 QY 1 MAPEINTKLTVPVHSAATGGERAYVTFLLAGTDYVGVVGLAKGLRKAASKYPLVAVLP 60  
 DB 2 mgpnvsse---kkalaaakarrayvtflagnqdyvkvglakglrrvrsayplvavlp 57  
 QY 61 DYPEDHRKQLVDOGCVVVKEIEPVPPENOTEFAMAYVINYSLRIWEFVEYNKMIYLDG 120  
 DB 58 dvpgghrrklveqgvvrelqpypesqcfamayvinykslriwefveymvylida 117  
 QY 121 DIOQVFNIDHFLDPLNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTWPPEAKLGPKP 180  
 DB 118 diqvfdnldihldfplndnyfayvmdcfccktwgthqyqgycqcpdevaweprelgppp 177  
 QY 181 PLYFNAGMFVEPNLSTYHNLLETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLYLAM 240  
 DB 178 plyfnagmfvepnlstytakdlldalvtvptpfaeqdflnmfreykpiipvnvnlvlam 237  
 QY 241 LWRHPENIELDQKVHVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWMDIYNDESLDYKN 300  
 DB 238 lwrhpenveldkvkvhycaagkpwrtgkeenmeredikmlvkrwmdiyndesldyke 297  
 QY 301 VVIGDSHKKQOOLQOFTIEALSEAGALQYVKAPSAA 335  
 DB 298 ----eednadeasqpmrtalaeagavkyfpapsaa 328

## RESULT 8

AAE05241  
 ID AAE05241 standard; Protein; 328 AA.

AC AAE05241;

DT 12-SEP-2001 (first entry)

XX Rice water stress-induced protein.

XX Rice; glycogenin; water stress-induced protein; starch biosynthesis;  
 KW glycogen; transgenic plant; plant breeding.

OS Oryza sativa.

XX US6255114-B1.

XX 03-JUL-2001.

PF 06-MAY-1998; 98US-0073297.

XX 07-MAY-1997; 97US-0852615.

PA (DUPO ) DU PONT DE NEMOURS & CO E I.

PI Lightner JE, Everard JD;

XX WPI; 2001-432051/46.

XX New glycogenin polynucleotides and polypeptides or water stress  
 PT proteins, useful for encoding homologous glycogenins and water stress  
 PT proteins from the same or other plant species -

XX Disclosure; Fig 2; 33pp; English.

XX The present sequence is water stress-induced protein from rice.  
 CC The present invention relates to plant glycogenin or water stress  
 CC proteins and their corresponding nucleic acid sequences. Glycogenin  
 CC is a self-glucosylating protein involved in the synthesis of glycogen  
 CC which plays a major role in starch biosynthesis. Over-expression  
 CC or reduction of glycogenin gene expression alters the branch chain  
 CC distribution of starch. Nucleic acid sequences of the present invention  
 CC are used to isolate cDNAs and genes encoding homologous glycogenins and  
 CC water stress proteins from the same or other plant species, or to create  
 CC transgenic plants in which an instant glycogenin or water stress protein





PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 25-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.



```
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160788.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 67.3%; Score 1222.5; DB 21; Length 279;
Best Local Similarity 78.5%; Pred. No. 7.3e-123;
Matches 219; Conservative 28; Mismatches 31; Indels 1; Gaps 1;

QY 58 VLPDVPEDHRKQLDQGVVKEIEPVYPPEQTEFAMAYVINYSKLRIFEVEYKMIY 117
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 mlpdpveehrilvdqgcivrieipvpyppeqtcqfamaayvinykskrlrwkfvayskmiy 60

QY 118 LDGDIQVFNIDHLDLPNGQFYAVMDCFCCKTWSHSPQYKIGYCQCQCPKVTWPEAKLG 177
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 ldgdiqvyenidhldfpdgylyavmdcfcektwshpqqykirycqcpdkvqpkaelg 120

QY 178 KPPLIFYNAGMFVTEPNLSTYHNLETVKIVPPTLFAEQDFLNMVFKDIYKPIPPVYNLV 237
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 eppalyfnagmflyepnalteydelrlrltkitpftptpfaeqdfnlmnyfkkiykpilpynlv 180

QY 238 LAMLRHRHPNTELDQVKVWHYCAAGAPWRFTGEEENMDREDIKMLVKKWMDIYNDESLD 297
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 lamlrwrpenvelqkvkvvhyaagsgkpwrytgkeanmeredikmlvkkwmdlyddesld 240

QY 298 YKN-VVIGDSHKKQOTLQOQFTEALSEAGALQYVKAPSAA 335
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 241 ykkpvtvtdtevdlnkpkftalteagrlnyvtapsaa 279

RESULT 11
AAB72462
ID AAB72462 standard; Protein; 117 AA.
XX AC AAB72462;
XX XX
XX DT 08-MAY-2001 (first entry)
XX DE Partial WSIP.
XX KW UGGT; UDP-glucose:glycoprotein glucosyltransferase; enzyme; WSIP; rice.
XX OS Oryza sativa.
```

```
XX WO200112845-A1.
XX PN
XX PD
XX PF 22-FEB-2001.
XX PR 27-JUL-2000; 2000WO-CA00883.
XX PR 18-AUG-1999; 99US-0376330.
XX PR (CANA ) NAT RES COUNCIL CANADA.
XX PI Tessier DC, Dignard D, Bergeron J, Thomas DY;
XX WPI; 2001-218358/22.
XX PT Determining the effect of a test sample on UDP-glucose:glycoprotein
XX glucosyltransferase (UGGT), useful for measuring UGGT activity,
XX PT comprises exposing an acceptor substrate for UGGT to a labeled donor in
XX PT the presence of UGGT.
XX PS Disclosure; Fig 10; 95pp; English.
XX CC The present invention relates to a method for determining the effect of a
XX CC test sample on UDP-glucose:glycoprotein glucosyltransferase (UGGT)
XX CC activity. The method comprises exposing an acceptor substrate for UGGT to
XX CC a labelled donor in the presence of the test sample and UGGT. The method
XX CC is useful for determining UGGT activity. In particular, the method is
XX CC useful in glucosyltransferase assay and kinetics measurement for
XX CC determining UGGT activity. UGGT is a soluble enzyme of the endoplasmic
XX CC reticulum which catalyses the addition of a glucose residue onto
XX CC asparagine-linked oligosaccharides, which are present on incorrectly
XX CC folded glycoproteins. The present sequence was used in a sequence
XX CC homology comparison with rat UGGT (see AAF60732 and AAB72436) which was
XX CC used in the method of the present invention.
XX SQ Sequence 117 AA;

Query Match 27.3%; Score 496; DB 22; Length 117;
Best Local Similarity 72.6%; Pred. No. 3.2e-45;
Matches 85; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

QY 109 FVEYNKMYLDGDIQVFNIDHLDLPNGQFYAVMDCFCCKTWSHSPQYKIGYCQCQCPDK 168
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 fveyermvyladaiqvfndhldkdgafyavkdcfcktwshpqqydygycqrpde 60

QY 169 VTWPEAKLGPKPLIFYNAGMFVTEPNLSTYHNLETVKIVPPTLFAEQDFLNMVFKD 225
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 vawperelgppppliyfnagmfvhepglgtaktldalvtvptpfaeqdfnlmfire 117

RESULT 12
AAB72556
ID AAB72556 standard; Protein; 346 AA.
XX AC AAB72556;
XX XX
XX DT 09-FEB-1999 (first entry)
XX DE Corn glycogenin clone cc3.mn0001.f7 protein.
XX KW Glycogenin; water stress protein; starch biosynthesis; corn; rice;
XX KW wheat; plant; granules; branch chain; endosperm.
XX OS Zea mays.
XX XX
XX FH Key Location/Qualifiers
XX FT Protein 1..346
XX FT /note= "Partial sequence, no start codon given"
XX PN WO9850553-A1.
XX PD 12-NOV-1998.
```







```

QY 219 LMYFKDIYKPIPPVYNLVNLAHLWRHPN-----IELDQVKVVHYCAAGAKPWR 267
Db 434 lnevftwvhr-iphmhf-lkhfwlgdeddakarrrkktelfgaepvlyvlhy--lgmkpwl 489
QY 268 FTGEEENMDREDI-----KMLVKKKWDIYNDESIDYKNVVIGDSHKKKQOTLQOFIEALS 321
Db 490 cyrdydcnfnsdifevfatdiahrkwmvhd-----ampqelhqfcylrs 534
QY 322 EAGA-LQYVKAPSAA 335
Db 535 kqkaqleydrrgaea 549

```

Search completed: July 2, 2002, 12:08:27  
Job time: 164 sec

